TS-79-106

Modelling Work Trip Distribution Patterns in Urban Ontario With Census Data

Ontario

Ministry of
Transportation and
Communications

Research and
Development
Division

Modelling Work Trip Distribution Patterns in Urban Ontario With Census Data

B. G. Hutchinson
D. P. Smith

Department of Civil Engineering
University of Waterloo

Final Report
OJT and CRT
Project No. 8510
Systems Research and Development Branch
TS-79-106

February, 1979

Ontario

Ministry of Transportation and Communications

Research and Development Division

TABLE OF CONTENTS

LIST OF FIGURES

Page No.

Page No.

LIST OF TABLES

Page No.

Page No.

ACKNOWLEDGMENTS

The data analyses on which this report is based were conducted by W. Arnold, M. Dias and B. Underhill undergraduate students at the University of Waterloo. These three students were employed as work term students under the co-operative education program conducted by the University of Waterloo.

Several members of the staff of the Ministry of Transportation and Communications provided important advice throughout the research project and these are V. C. Ma, P. Dalton and J. Saunders.

Mr. Louis Shallal of the Regional Municipality of Ottawa-Carleton and Mr. Gerry Thompson of the Regional Municipality of Waterloo kindly provided information on road network speeds for their respective municipalities.

Mrs. I. Steffler typed the report.

SUMMARY

This report describes some detailed trip distribution analyses of the 1971 census journey to work data for the fifteen Ontario census areas other than the Toronto CMA. Earlier studies of the commuting patterns in these fifteen Ontario urban areas had demonstrated that the commonly used gravity model is capable of explaining only the broad patterns of interaction between home and work and that significant over - and under - estimation residuals existed between the observed and simulated trip interchanges. A number of potential opportunities for improving the gravity model were identified in the earlier study and the principal objective of the current study was to explore the impacts of these modifications to the gravity model structure.

The cell entries in observed home to work linkages matrices reflect two types of effect and these are the rate of interaction between a pair of zones and the absolute magnitudes of the labour force and the employment in those zones. The essential hypothesis of the gravity model is that the rate of interaction between zones is a function of the spatial separation of zones. In order to understand better the spatial interaction patterns in the Ontario census areas it is essential to extract from home to work linkages matrices the pure spatial interaction effects that are independent of zone size effects.

Bi-proportional balancing techniques have been used to transform the home to work linkages matrices so that the entries in each row and in each column sum to a constant magnitude. The effect of this operation is to produce matrices in which the entries represent the interaction magnitudes that would occur between zones of equal size, or the pure interaction effects between home and work zones. This balancing process is straight forward in that the observed matrix row entries are each scaled by the same proportion so that the row totals are all equal to an arbitrary constant. The columns of this adjusted matrix are then summed and scaled so that the column totals are all equal to the arbitrary constant. This row by column adjustment process is continued iteratively until convergence is obtained with each row total and each column total being equal to the arbitrary constant. If the arbitrary constant were equal to one then the cell entries in the adjusted matrix could be interpreted as the unit probabilities of interaction between zones.

A clustering technique was then used to group zones on the basis of similarities in their destination vectors of the bi-proportionally adjusted matrix. The technique used is Ward's method which uses a measure of the Euclidean distance between destination vectors to cluster census tracts. The procedure begins with the n census tracts in a census area and groups them into a sequence of clusters until all census tracts are merged into a single cluster. The heterogeneity of the census

tracts within a cluster may be represented by the within cluster sum of squared differences in the destination vectors of the census tracts. The sum of the sum of squared differences over all clusters provides an indication of the overall heterogeneity in the commuting patterns from the residential zones at each stage of the clustering process. The error sum of squares increased slowly during the first stages of the clustering process and began to increase sharply when residential zones with quite different destination vectors are forced into the same cluster. The hierarchical structure of the clustered residential zones in a particular census area is conveniently represented graphically by a dendrogram. The dendrograms show the sequence in which zones cluster along with the error sums of squares magnitudes at which they cluster.

Dendrograms are presented for the fifteen Ontario census areas and commuting sub-regions are identified in each area on the basis of the cluster analysis. A detailed review of these commuting sub-regions suggests that there are five broad categories of determinants of these regions and these are: (i) multi-community composition of a census area; (ii) topographic and man-made features; (iii) the time sequence of development of an area; (iv) socio-economic factors; and (v) the domination of large employment concentrations.

Thunder Bay, Kitchener, St. Catharines and Hamilton census areas provide examples of the impact of municipal boundaries on commutersheds. In almost all cases of clustering at the four or five cluster level the cluster boundaries are coincident with municipal boundaries. The best examples of the impacts on commuting patterns of topographic features are provided by London, Hamilton and Ottawa. Timing of development influences are noted in almost all of the census areas particularly those that grew rapidly during the 1966-1971 period. The strongest determinant of commuting patterns appears to be the socio-economic characteristics of the households and in many cases the other determinants mentioned above co-vary with the socio-economic influences. More formal analyses of the determinants of commuting patterns could be undertaken by performing cluster analyses of household characteristics and exploring the extent to which these cluster boundaries are coincident with the commutershed boundaries.

Previous studies of the gravity model have shown that the commonly used goodness-of-fit statistics are inadequate for assessing the extent to which gravity model calculated trip matrices agree with the observed trip matrices. The effectiveness of a range of statistics is explored by a series of simulation experiments for eight of the larger Ontario census areas. Simulated trip matrices were calculated from the observed trip matrices by multiplying each of the observed cell entries by a randomly generated error with a specified range. A sequence of simulated trip matrices were prepared for six error ranges from ten percent to one hundred and fifty percent and magnitudes of the goodness-of-fit statistics were calculated for each of the six error ranges.

Detailed analyses of the changes in the magnitudes of the goodness-of-fit statistics with increasing induced error magnitude have shown that the so-called phi statistic of information theory is the most appropriate statistic to use. This statistic increased linearly with increasing error magnitude and resulted in a consistent ordering of the eight census areas across all census areas. While the phi statistic magnitude increases with increasing numbers of trips at a constant error magnitude a normalized phi magnitude was found to be roughly constant across all census area sizes. Accordingly the phi statistic divided by the number of trip interchanges in an area has been adopted as the most appropriate goodness of fit measure for evaluating changes to the gravity model structure.

A large number of variations in the gravity model have been tested for the Ontario census areas. A very flexible gravity model computer program has been developed which allows a wide range of gravity model structures to be tested and very exhaustive goodness-of-fit statistics to be calculated. In most cases gravity models have been calibrated for the eight largest Ontario census areas ranging in size from Thunder Bay to Ottawa.

Production constrained, attraction constrained and doubly constrained versions of the gravity model have been calibrated using road network distances in kilometres as the surrogate for the generalized travel costs between zones. This basic set of models was calibrated in order to provide the datum against which changes in the gravity model might be assessed. The calibration criterion used for all model types is the minimization of the sum of the absolute differences in the ordinates of the observed and simulated trip length frequency distributions. In most cases this criterion also satisfied the more normally used criterion of equality of observed and simulated area-wide mean trip lengths.

Using the normalized phi statistic as the goodness of fit measure of the calibrated models suggested that these models produced trip tables that were of the same statistical quality as those produced by the random error generator with an error range of 75 to 100 percent. A detailed comparison of the model types showed that the doubly constrained model was consistently superior to the production and attraction constrained versions of the gravity model with the attraction constrained version having the worst goodness of fit. These analyses also showed that the goodness of fit deteriorated significantly with increasing urban area size. While the doubly constrained model provided the best goodness of fit there are extreme difficulties in providing any reasonable interpretation of the travel deterrence parameter and balancing factor magnitudes.

Doubly constrained models using network travel times rather than network distances have been calibrated for the Kitchener, Hamilton and Ottawa census areas and compared with those developed in terms of network distances. The phi-statistics showed that these models employing travel times were marginally inferior to those calibrated in terms of network distances. Analyses of the zone-specific goodness of fit statistics showed that in the Ottawa CMA the time-based models

improved the goodness-of-fit of some of the outlying residential census tracts but that this was not a general trend across all census areas.

Inter-zonal generalized travel costs were also created from various combinations of network distances and network times for the Kitchener CMA. Analyses of the goodness of fit statistics showed that doubly constrained models calibrated using these generalized cost functions are superior to those using travel times but marginally inferior to the models calibrated using network distances.

Detailed comparisons of the trip interchange residuals for the Ottawa CMA and a partitioning of the phi-statistic into over - and under - estimation residuals by intra-zonal and inter-zonal trip interchanges showed that the major source of error is the under-estimation of trip interchange magnitudes and that the superior behaviour of the doubly constrained model is due to a reduction in the under-estimation residuals. However this is not a completely unbiased interpretation of the residuals since earlier simulation studies had shown that the phi-statistic is more sensitive to under-estimation errors than to over-estimation errors.

A number of stratified gravity models have been calibrated using all three basic forms of the gravity model. The three groups of stratified gravity models calibrated are: (i) multi-parameter gravity models with separate travel deterrence function parameters estimated for each census tract; (ii) sub-region specific models in which separate travel deterrence parameter magnitudes are calibrated for from four to eight commuting sub-regions depending on the specific area; and (iii) a set of models calibrated for separate socio-economic groups.

The multi-parameter production constrained models all exhibited superior goodness of fit characteristics to the single parameter production constrained models but were not as good as the single parameter doubly constrained models. The largest improvements occurred in the multi-community census areas of Kitchener and St. Catharines. The multi-parameter attraction constrained models were inferior to the single parameter versions except for Thunder Bay, Kitchener and St. Catharines. The multi-parameter doubly constrained models all exhibited superior goodness of fit characteristics to the single parameter versions.

The cluster analyses of the census areas formed the basis for the identification of the calibration sub-regions where these ranged from three in Thunder Bay and Kitchener to eight in the Ottawa census area. The sub-region specific production constrained models are superior to the single parameter versions but are inferior to both the single and multi-parameter production constrained models. An analysis of the sub-region specific doubly constrained models showed that they performed marginally better than the single parameter version but are inferior to the multi-parameter models.

The most useful models stratified by socio-economic characteristics are those developed for non - car owners and car - owners. Detailed studies for the Kitchener area showed that doubly constrained versions of these stratified models perform at a superior level to the single parameter doubly constrained models.

Trip generation rates have been established for each of the census areas for households stratified by tenancy status and by period of residence. Owner occupied dwelling units have significantly higher numbers in the labour force than the rented dwelling units. The number of home to work linkages from households with periods of residency greater than five years is larger than for households with shorter periods of residency.

Detailed regression analyses have been conducted in order to establish transit captive and non-captive rates for all census areas. Statistically significant prediction equations have been developed in terms of the dwelling unit composition of census tracts and there are significant differences in the rates between census areas. Transit captive generation rates from single attached dwelling units are from two to three times greater than from single detached dwelling units. Some of the variation in generation rates is due to the variation in the labour force participation rates between census areas and a second set of regression equations has been developed from the data factored to a constant labour force participation rate across all census areas. Pooled regression equations were also developed for the entire set of census areas. The pooled equations show that for average labour force participation rates the captive generation rate is 2.3 times larger for attached dwelling units than for detached dwelling units and that the non-captive rate is about 3 times larger for detached dwelling units than for attached dwelling units. Significant relationships between captive and non-captive trip attraction and employment sector type could not be established.

INTRODUCTION

The practical importance of the journey to work data collected in the 1971 Census of Canada has been recognized for many years by the Ministry of Transportation and Communications of Ontario. This organization has been responsible for the use of census information in a number of transport model building and research projects conducted both within the Ministry and by external agencies. This report describes the results of a research project sponsored at the University of Waterloo by the Ministry of Transportation and Communications which has been concerned with a more detailed analysis of the spatial interaction patterns in Ontario census areas observed in the 1971 census. This project represents an extension to an earlier research project conducted during 1976-1977 which was also concerned with the analysis of the census data.

The primary objective of these two research projects was to explore the use of the census journey to work data in the development of generalized transport models that might be used to synthesize future work trip demands without the collection of comprehensive special purpose surveys. A second objective of these studies was to develop a better understanding of spatial interaction patterns across a wide range of urban areas and the factors that influence these interaction patterns.

The research project completed in January, 1977 showed that regression equations with very high statistical qualities could be developed from the census data for estimating the amount of labour force in each census tract from a knowledge of the dwelling unit composition of census tracts. Various forms of the gravity model were

calibrated using the census journey to work data. While these calibrated models explained much of the observed spatial interaction between census tracts large positive and negative trip interchange residuals were observed in all urban areas. It was suggested that a variety of factors influenced these residuals and these included the timing of development, socio-economic factors and the nature of the deterrence function embodied in the gravity model.

In an attempt to develop better trip distribution models a number of specific objectives were set for the research project described in this report and these included:

1. To use bi-proportional matrix balancing techniques to calculate normalized trip linkage magnitudes between each pair of origin-destination zones in order to separate interaction effects from zone size effects

2. To use clustering techniques to identify groups of census tracts with similar interaction behaviour in order that factors influencing spatial interaction may be isolated and incorporated in the gravity model structures

3. To develop a more flexible gravity model computer program that allows alternative model types incorporating different deterrence functions to be calibrated and model errors to be examined in detail

4. To examine various statistical criteria that may be used to assess the goodness of fit of trip distribution models

5. To estimate the parameter magnitudes of alternative forms of the gravity model which have been calibrated by geographic sub-region, socio-economic group and which incorporate various travel deterrence functions and travel deterrence measures.

CHAPTER 1

THE CENSUS JOURNEY TO WORK DATA

The characteristics of the 1971 journey to work data and the urban road networks which have been coded to supplement the spatial interaction data are described in detail in ref. [1]. However, it is worthwhile recalling the general characteristics of the data base used in the research project described in this report.

The basic spatial unit to which the 1971 census journey to work data are coded is the census tract where the average residential census tract contains about 5,000 people. In 1971 the number of census tracts varied from 14 in Guelph (population = 62,660) to 452 in the Toronto CMA (population = 2,628,000). The coded journey to work data represent approximately an 11 percent sample of the journey to work by households. The contiguous urbanized portions of a census area are covered by the official census tract system and the analyses described in this report deal with the interactions between official census tracts. Home to work linkages also occur between official census tracts and the surrounding census districts and with other census areas but these interactions are excluded from the analyses. In most of the Ontario census areas the home to work linkages between official census tracts within a census area account for about 80 percent of the total interaction.

Road networks have been coded at roughly the arterial/sub-arterial level in terms of road distances. In addition, link travel speed data have been added for selected urban areas.

1.1 The Ontario Census Areas

At the 1971 census there were 16 census areas in Ontario which ranged in population size from Guelph to Toronto. This report is concerned with the journey to work in 15 of these census areas with the Toronto CMA being excluded from the analysis. Table 1 shows the populations of the 15 census areas in 1971 along with the number of census tracts in each area. This table shows that six of the census areas had populations of less than 100,000, three had populations between 100,000 and 200,000, four had populations between 200,000 and 300,000 and two had populations of 500,000 or greater.

In addition to the variations in population size the census areas also varied widely in economic base, spatial character and geographic setting. In 1971 three broad industry groups accounted for about 70 percent of the jobs in the Ontario census areas and these were manufacturing, trade and services. The principal differences between the economic bases of the census areas were in the primary (mining), manufacturing, transportation and public administration sectors. In Sudbury, 26.5 percent of the jobs were in mining while Brantford, Oshawa and Kitchener had strong manufacturing industry bases. Employment in the transportation sector was well above the average in Thunder Bay while 16.3 percent of the labour force in Ottawa was in public administration.

These variations in the employment bases of the Ontario census areas were reflected in the spatial distributions of employment. The geographic settings of the census areas also have a strong influence on their spatial characters. Eight of the Ontario census areas have heavy concentrations of employment in two or three census tracts while in four

TABLE 1. 1971 Census of Canada Census Metropolitan Areas and Census Agglomerations for Ontario

C(M)A Number		1971 Population Total C(M)A	1971 Population Official CTs	Number of CTs	Number of CDs
1	Guelph	62,660	60,087	14	19
2	Peterborough	63,530	58,111	16	10
3	Sarnia	78,445	60,417	18	6
4	Brantford	80,285	70,904	16	16
5	Sault Ste. Marie	81,270	80,332	19	2
6	Kingston	85,875	59,047	18	8
7	Thunder Bay	112,095	108,411	25	2
8	Oshawa	120,320	116,911	22	10
9	Sudbury	155,425	94,624	28	5
10	Kitchener	226,850	212,819	45	21
11	Windsor	258,645	211,494	56	5
12	London	286,010	223,222	59	15
13	St. Catharines	303,430	221,282	53	11
14	Hamilton	498,525	437,554	109	19
15	Ottawa	602,510	506,316	120	19

of the census areas residential areas are located away from the job centres creating longer average trip lengths. In several census areas these effects combine to produce long average trip lengths that are unrelated simply to population size.

1.2 Previous Trip Generation Analyses

Ref. [1] describes in detail the regression analyses conducted in the 1976-1977 research project which were directed towards the development of equations for predicting the labour force in each census tract given a knowledge of the amount of residential activity in a tract. It must be remembered that in this report the term trip generation really means the number in the labour force in each zone rather than the number of trips. The best prediction equations were developed in terms of the dwelling unit composition of census tracts and these prediction equations are summarized in Table 2. This table shows that the equations are of very high statistical quality and that the partial regression coefficients are quite stable across urban areas. The consistency and explanatory power of these equations would suggest that work trip production studies be standardized in terms of dwelling unit composition. Trip attraction analyses could not be performed with the aggregate data since independent measures of employment by census tract were not available.

1.3 Previous Trip Distribution Analyses

Three forms of the tradional gravity model were calibrated from the census data and these were a production constrained model incorporating an inverse power deterrence function and a production constrained model with attraction trip end balancing incorporating inverse power and

TABLE 2. Census Tract Labour Force Prediction Equations Using Dwelling Unit Composition

Census Area	a	b_1	b_2	R^2
Guelph	35	1.571	1.172	0.96
Peterborough	0	1.544	0.909	0.98
Sarnia	-63	1.526	0.983	0.98
Brantford	26	1.566	0.925	0.99
Sault Ste. Marie	-51	1.580	1.315	0.99
Kingston	-28	1.669	1.181	0.99
Thunder Bay	45	1.438	0.919	0.96
Oshawa	-26	1.524	1.115	0.96
Sudbury	24	1.555	1.394	0.97
Kitchener	-87	1.762	1.310	0.98
Windsor	-77	1.558	0.919	0.97
London	197	1.474	1.016	0.97
St. Catharines	98	1.435	0.798	0.99
Hamilton	13	1.516	1.083	0.97
Ottawa	51	1.660	1.212	0.96

CT Labour Force = a + b_1. CT Single Detached Dwelling Units

+ b_2. CT Single Attached Dwelling Units and Apartments

negative exponential deterrence functions. While there were some differences in the overall performance of the various model types all of the calibrated models produced large residuals between simulated and observed trip interchange magnitudes.

Figure 1 illustrates the spatial distributions of under- and over-estimation residuals for the production constrained version of the gravity model in London. The trip interchange residuals plotted in Figure 1 are equal to the estimated trip interchange magnitudes minus the observed trip interchange magnitudes. Inspection of the residuals presented in Figure 1 shows that most of the under-estimation residuals are associated with non-CBD employment locations while almost all of the over-estimation residuals are associated with the CBD. Employment in London in the central census tract is very much higher than in the other census tracts and in the production constrained model trips tend to be diverted to the central zone from competing employment locations.

Figure 2 illustrates the spatial distributions of under- and over-estimation trip interchange residuals for the doubly constrained form of the gravity model calibrated for London. This diagram demonstrates that the strong over-estimation of trips to the central area is reduced significantly by the use of the attraction trip end balancing procedure. The balancing procedure reduced the attractivity of the central area and increased the attractivity of the non-CBD employment locations. Figure 2 illustrates that while trip end balancing reduced the majority of over-estimation residuals it had little impact on the under-estimation residuals. Similar effects to those illustrated in Figures 1 and 2 for London were observed for other Ontario census areas. The size of the

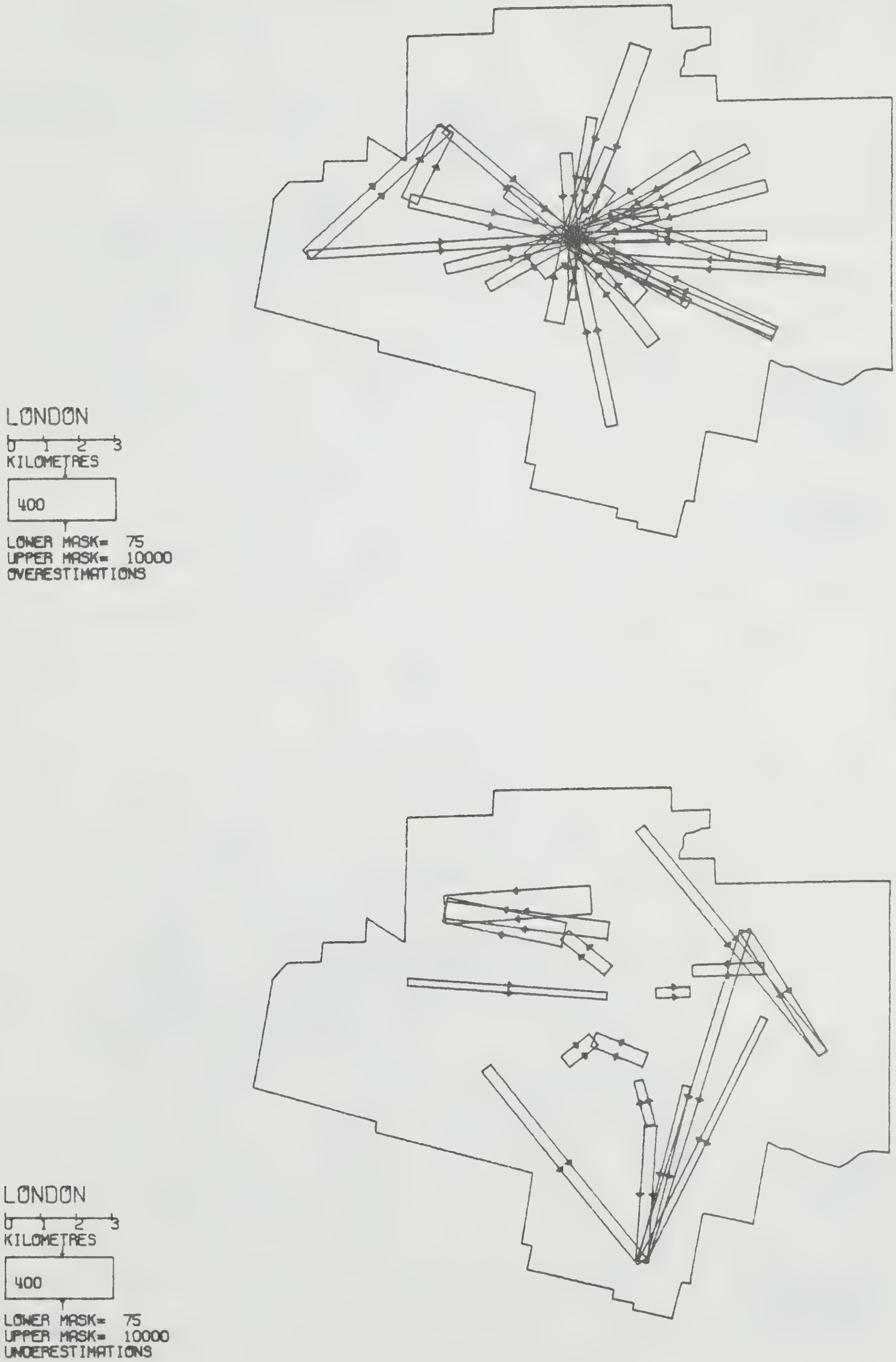
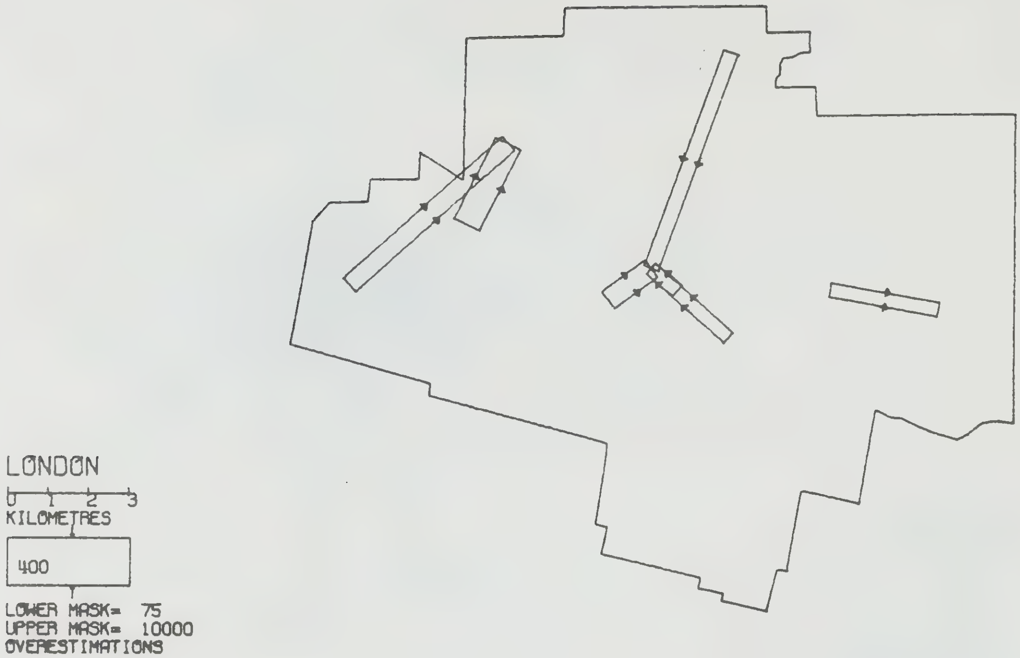


FIGURE 1. Production Constrained Gravity Model Trip Interchange Residuals for London



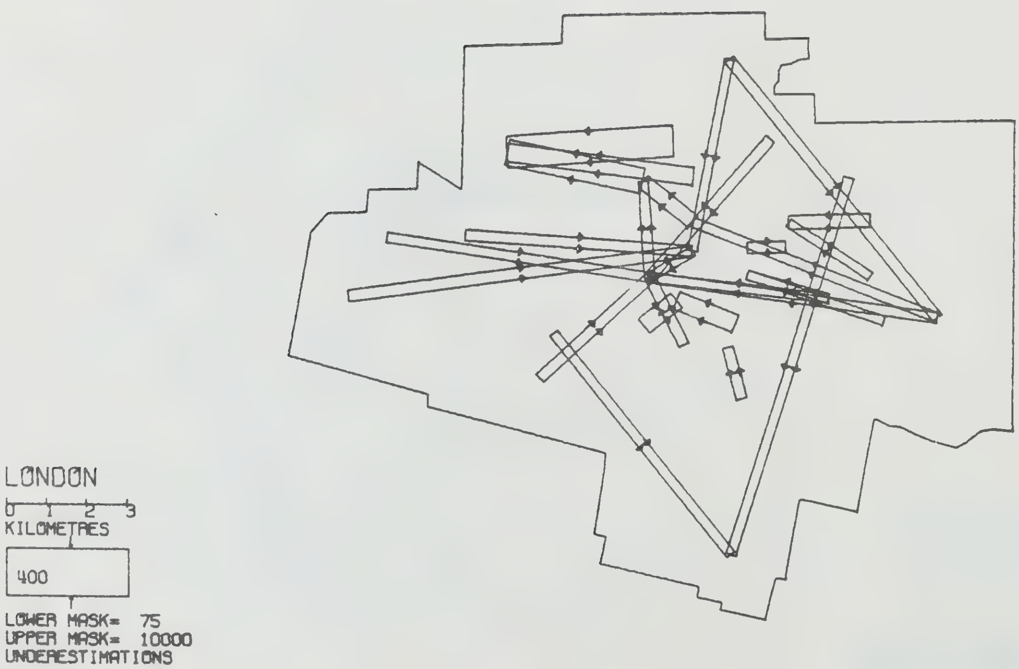


FIGURE 2. Doubly Constrained Gravity Model Trip Interchange Residuals for London

residuals indicated that significant modifications to the gravity model structure are required if spatial interaction magnitudes are to be estimated more realistically.

1.4 Additional Analysis Opportunities

In the earlier research project the gravity models were calibrated using the aggregate home to work matrix. This approach assumes implicitly that the interaction patterns observed in 1971 are influenced only by the spatial distributions of production and attraction trip ends and the spatial separation of zones. It has been suggested previously that this simple hypothesis is inadequate and that other factors have an important influence on home to work linkages.

The gravity models traditionally calibrated have a single deterrence function which implies that the response to travel time of residents in all parts of an urban area is similar. Figure 3 shows the home to work linkages greater than 250 between census tracts in the Kitchener census area. This diagram illustrates that commuting tends to be relatively self-contained within each of the four local municipalities and that the characteristic trip lengths tend to vary between areas. This suggests that one potential improvement to the gravity model might be to estimate behavioural parameters for the separate sub-regions of the area.

A second characteristic of urban areas that is not reflected in the traditional form of the gravity model are the differences in commuting behaviour between various socio-economic groups. Clearly the interaction spaces of people without access to a car for the journey to work will be restricted. Figure 4 shows the spatial distributions of productions and attractions for car-owners and non-car owners in the Kitchener census area.

FIGURE 3. Home to Work Linkages in Kitchener Census Area

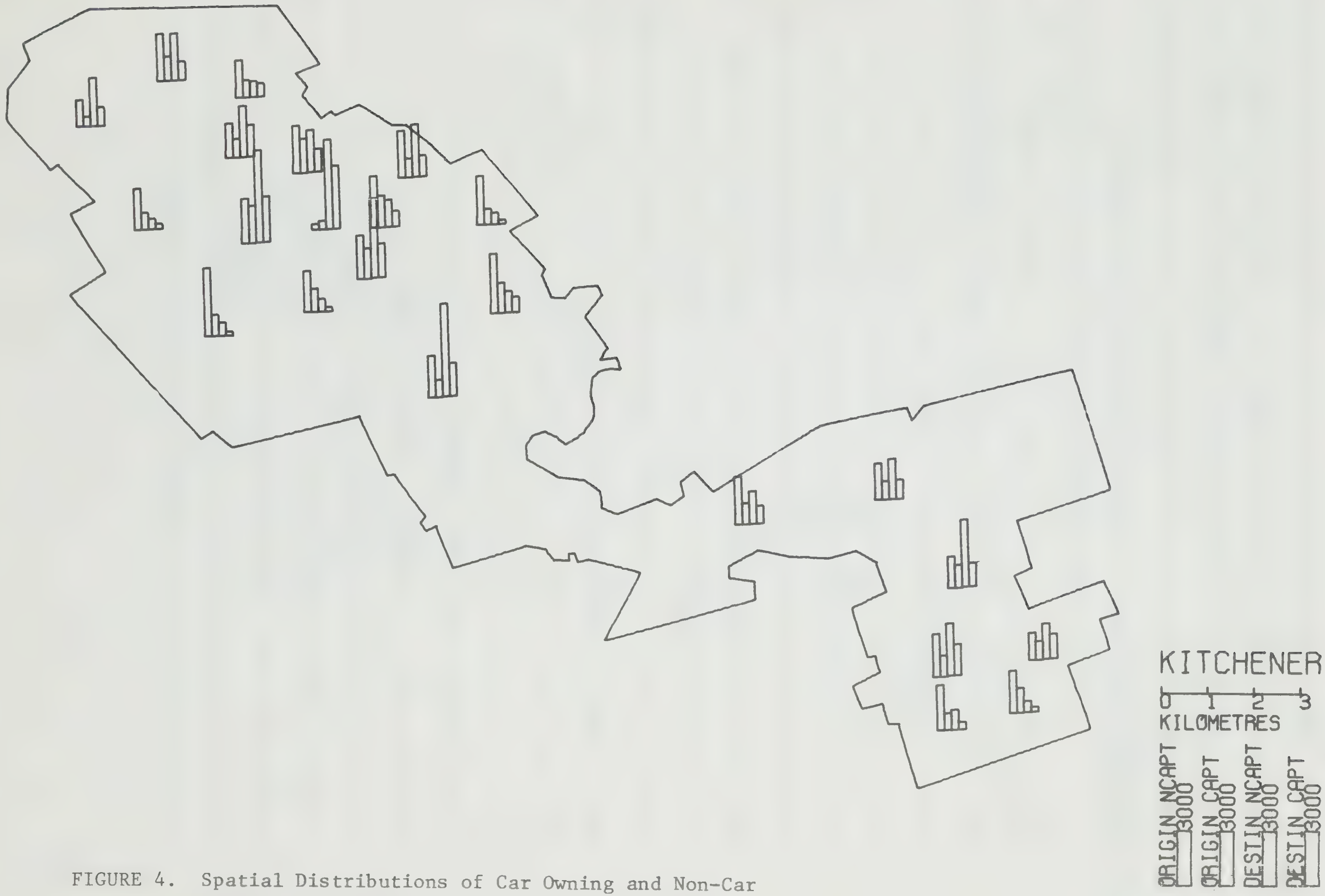


FIGURE 4. Spatial Distributions of Car Owning and Non-Car Owning Origins and Destinations in Kitchener Census Area

This diagram illustrates that households containing non-car owning members of the labour force and jobs attracting non-car owners are not distributed uniformly throughout the area but tend to be clustered around the central area of Kitchener. A second potential improvement to the gravity model structure is to stratify it by socio-economic group in order to reflect potential differences in commuting behaviour between groups.

A third factor influencing spatial interaction patterns is the staging of development of an urban area. The location decisions of individuals at a particular time will reflect the spatial distributions of household and employment opportunities that exist at that particular time. It would be expected that households which have located in recent times would tend to be concentrated on the periphery of the urbanized area where new residential sub-divisions are developing. In addition one would expect that these new residential areas would be linked strongly with those areas of new employment growth. These types of effects are illustrated in Figures 5 and 6 for the Kitchener census area. Figure 5 shows the home to work linkages for households which have been located for six years or more, that is, prior to 1966. In the Kitchener area most of these linkages are between households located in the middle and inner suburbs and the inner industrial areas with some linkages to jobs in the industrial park in the southern sector of Kitchener. Figure 6 illustrates that the spatial linkages of households which have been located for five years or less have a different character with the bulk of the trips originating from the outer residential areas. Many of these trips are to the University of Waterloo and the Kitchener CBD both of which grew rapidly during the 1966-1971 period. A third opportunity for improving the predictive powers of the gravity model

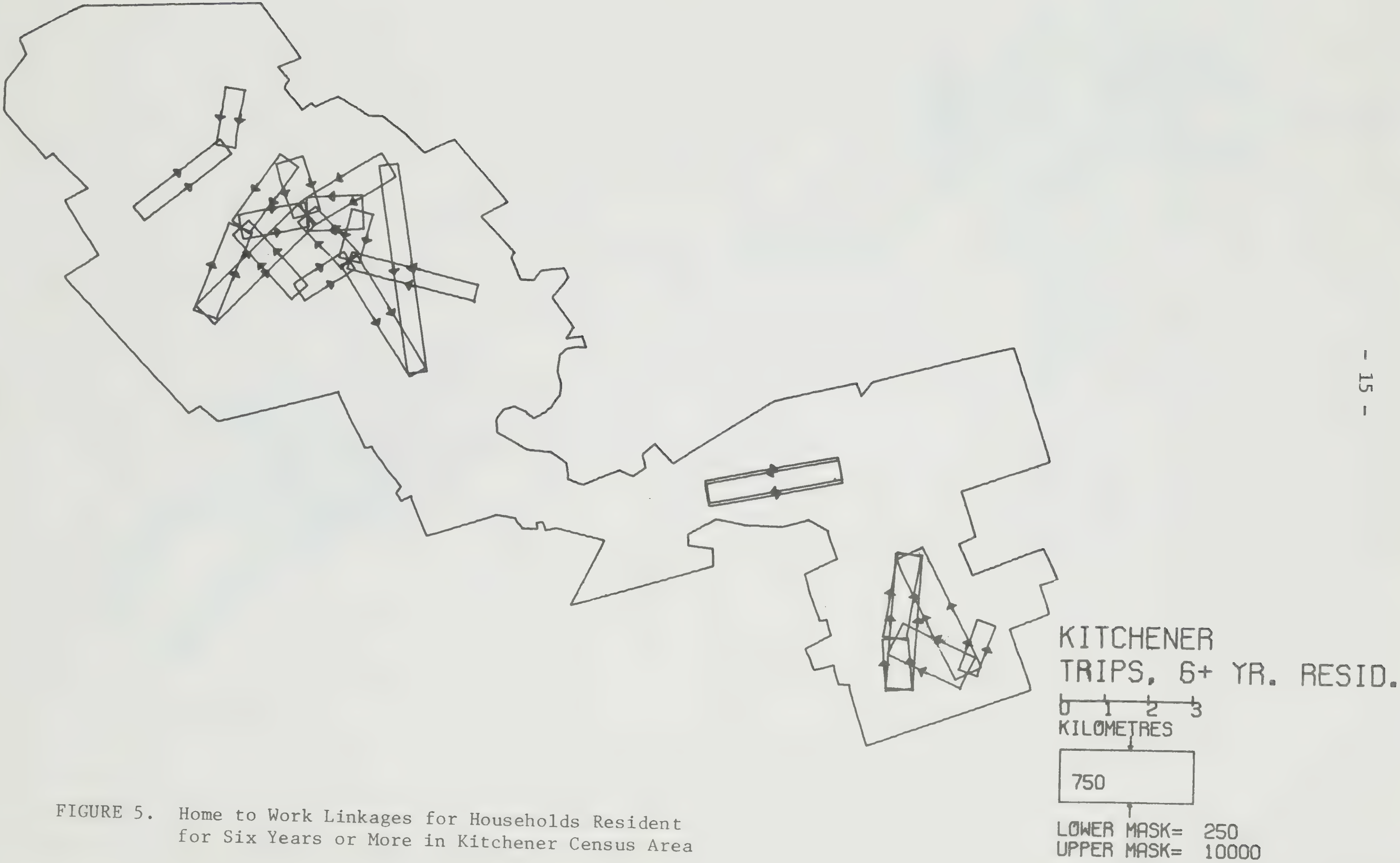


FIGURE 5. Home to Work Linkages for Households Resident for Six Years or More in Kitchener Census Area

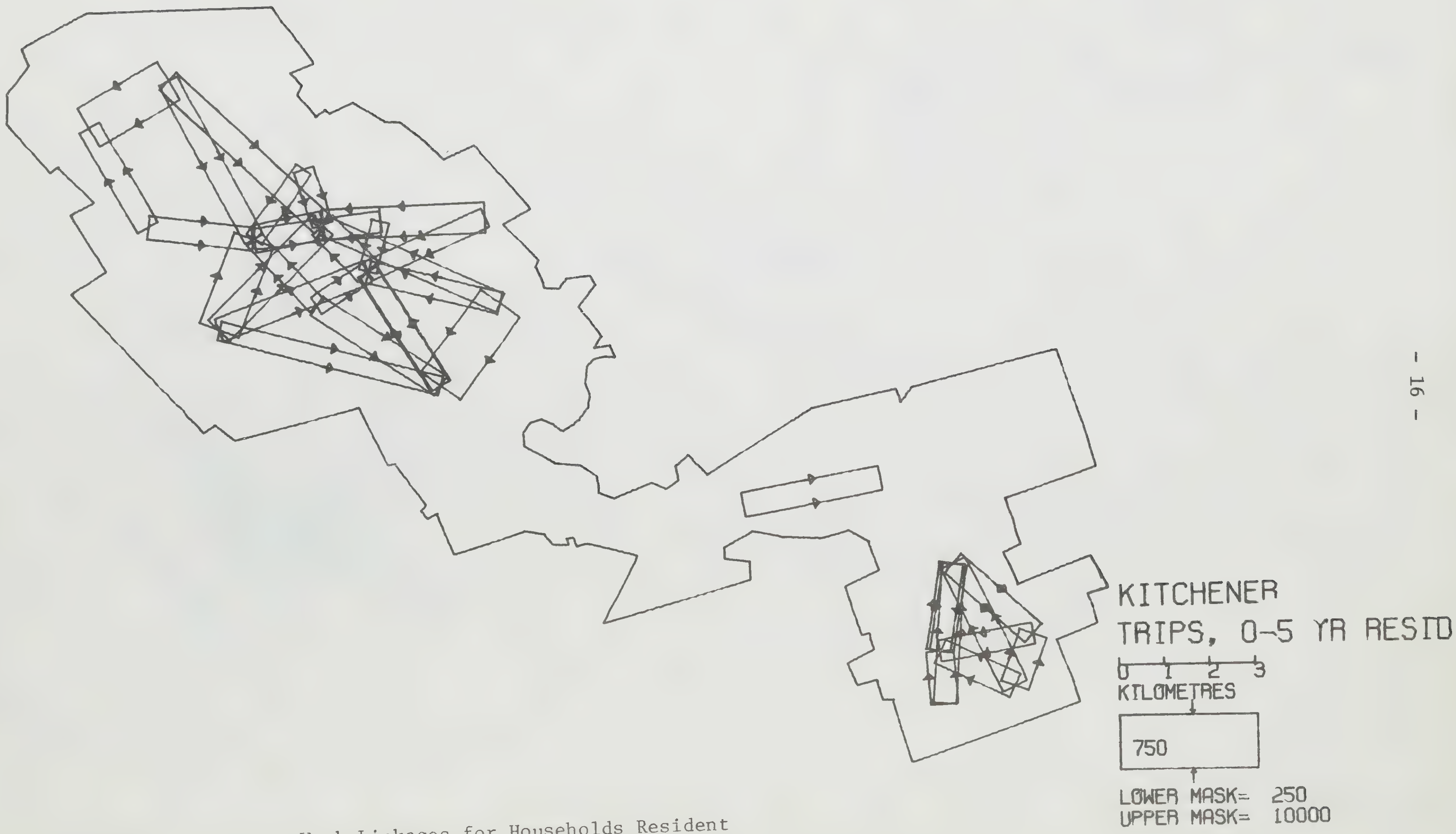


FIGURE 6. Home to Work Linkages for Households Resident for Five Years or Less in Kitchener Census Area

is to recognize that home to work linkages develop in a sequential way and not instantaneously as implied by the traditional gravity model. Clearly, each of the three types of effects mentioned previously are not mutually exclusive and there are some overlaps between behaviour, socio-economic characteristics and the timing of development.

CHAPTER 2

EMPIRICAL ANALYSES OF SPATIAL INTERACTION

The previous chapter has identified a number of opportunities for improving the ability of the gravity model to explain observed spatial interaction between home and work. The gravity model estimates spatial interaction in terms of the product of the production and attraction trip end magnitudes and the travel deterrence between zones. Some of the difficulties of improving the gravity model are related to this confounding of the zone size effects and the interaction effects.

The cell entries in observed trip matrices reflect these two types of effect, the rate of interaction between a pair of zones and the labour force and employment magnitudes of the zones. In order to understand better the spatial interaction patterns it is essential to extract from the home to work linkages matrices the pure spatial interaction effects that are independent of zone size effects. It is these rates of spatial interaction that trip distribution models attempt to explain in terms of the travel deterrence function.

This chapter describes some empirical studies of the journey to work matrices obtained in the 1971 census for the fifteen Ontario census areas. The analysis strategy consisted of two basic steps which are: (i) bi-proportional balancing of the trip matrices; and (ii) the clustering of zones in terms of similarities in destination characteristics.

The bi-proportional balancing techniques transform the home to work linkages matrices so that the entries in each row and in each column

sum to a constant magnitude. The effect of this operation is to produce matrices in which the entries represent the interaction magnitudes that would occur between zones of equal size, or the pure interaction effects between home and work zones. The clustering technique compares the destination vectors for each origin zone and groups zones together that have the most similar destination vectors. The clustering technique groups all zones into a single cluster using a step by step approach.

2.1 Bi-Proportional Matrix Balancing

It has been suggested above that bi-proportional balancing techniques may be used to remove the effects of zone size from trip matrices leaving only the information on the rate of spatial interaction between zones. If this is not done then zones with large amounts of employment will dominate any analyses and may mask important interaction effects between smaller zones. Bi-proportional matrix balancing techniques have been developed for use in a number of fields with much of the original work being conducted in connection with input-output tables [2]. The Furness method [3] of origin-destination table development is a closely related procedure.

The entries in bi-proportionally adjusted trip matrices represent the interactions that would occur if all the zonal productions and attractions in an urban area were equal. The row/column constant used in matrix balancing can have any arbitrary magnitude. For example, if it were equal to 1.0 then each cell entry could be interpreted as the probability of interaction between a known origin zone and a particular destination zone when all destination attractivities have the

same magnitude. Normalized trip matrices therefore reflect only spatial separation effects rather than both spatial separation and structural effects.

The calculation of bi-proportional matrices is straight forward. Trip matrix row entries are each scaled so that the row totals are all equal to the arbitrary constant. The columns of this new matrix are then summed and scaled so that the column totals are all equal to the arbitrary constant. This process is continued iteratively until convergence is obtained with all row and column totals being equal to the arbitrary constant.

It is useful to examine the properties of the transformed matrices in relation to the original trip matrix. The bi-proportional balancing techniques preserves the interaction rate structure of the original matrix but the cell entries in any row or column are not all adjusted by a factor with a constant magnitude. However it has been shown [2] that the products of the final row and column adjustment factor magnitudes have unique magnitudes for each cell. This is, the interaction magnitudes calculated by the bi-proportional balancing techniques are unique.

It is also useful to examine the changes in the trip length structure that might be expected from the transformation of the original trip matrix. The mean trip length of any origin or destination zone is calculated simply from the weighted sum of the trip distances to each of the other zones in the system. It has been noted previously that the bi-proportionalized matrix retains the interaction rate structure but not the interaction magnitude structure of the original matrix.

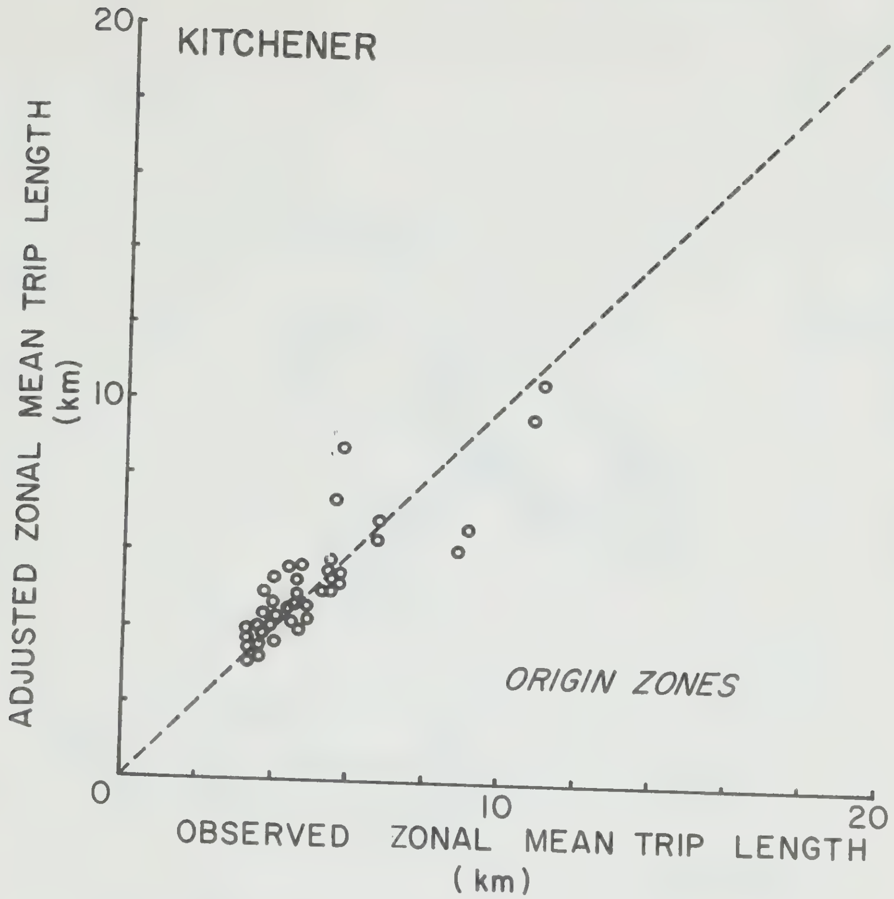
As a result the mean trip lengths for any origin or destination zone for the bi-proportionalized matrix may be expected to change since the row or column weight composition used will have altered. Rows which do not have strong concentrations of destinations at one or two zones may be expected to exhibit little change in the zone mean trip length. On the other hand zones with strong concentrations of destinations in a few zones might be expected to exhibit significant shifts in the mean trip length. The weights between zones with small trip end magnitudes will increase and a greater influence on the zone mean trip lengths of the bi-proportionalized matrices. The bi-proportional adjustment process will produce more uniform zone mean trip lengths than originally observed.

Table 3 summarizes the area-wide trip lengths for both the original trip matrix and the balanced matrix for the fifteen Ontario census areas. This table shows that in most cases the area-wide mean trip lengths of the balanced matrices are shorter than the observed. The largest shifts occur in those census areas with unusual trip linkage characteristics such as Sarnia, Kingston, Sudbury, Hamilton and Ottawa. In these areas a few employment zones tend to dominate the commuting structure.

Figures 7 to 9 provide more detailed comparisons of the adjusted and observed trip matrices in which the adjusted mean trip length for each zone is plotted against the observed mean trip length. Figures 7 and 8 provide comparisons for the origin zones in six census areas while Figure 9 shows comparisons for the destination zones of Hamilton and Ottawa.

TABLE 3. Comparison of Area-Wide Observed and Balanced Trip Matrix Mean Trip Lengths

	Observed Mean Trip Length km	Balanced Mean Trip Length km
Guelph	3.76	3.77
Peterborough	3.30	3.36
Sarnia	6.33	5.44
Brantford	4.20	3.71
Sault Ste. Marie	4.59	4.55
Kingston	5.68	4.32
Thunder Bay	5.37	5.65
Oshawa	4.10	4.10
Sudbury	8.63	7.09
Kitchener	4.80	5.03
Windsor	7.50	7.31
London	6.57	6.70
St. Catharines	6.50	6.23
Hamilton	8.05	7.10
Ottawa	7.63	6.88



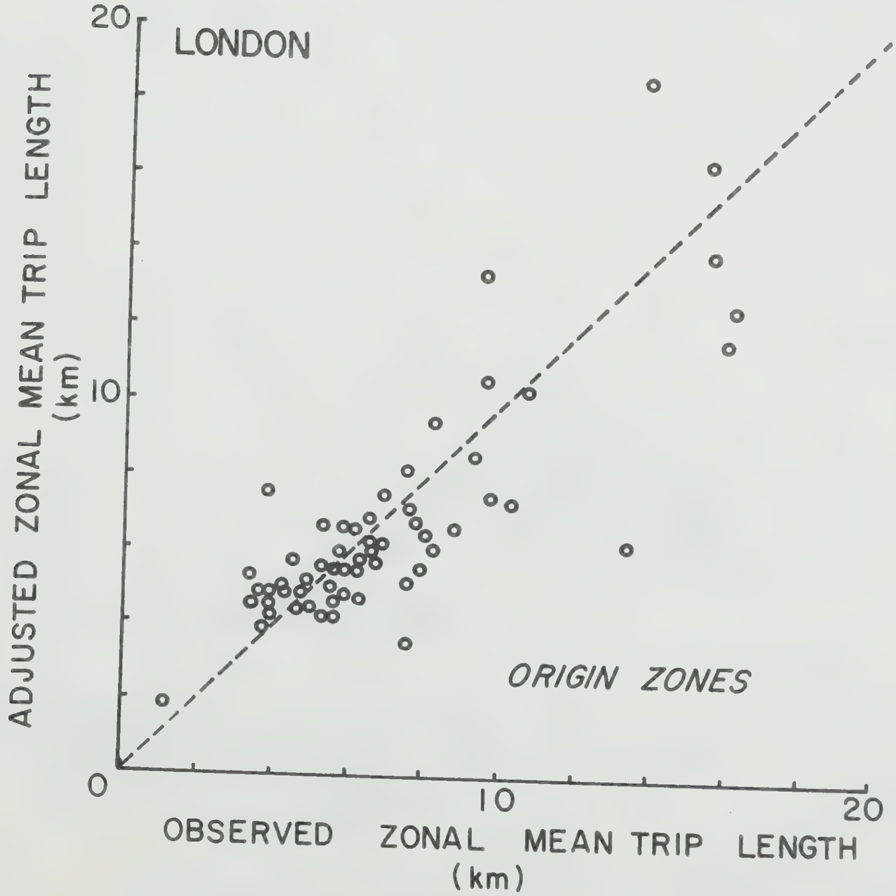
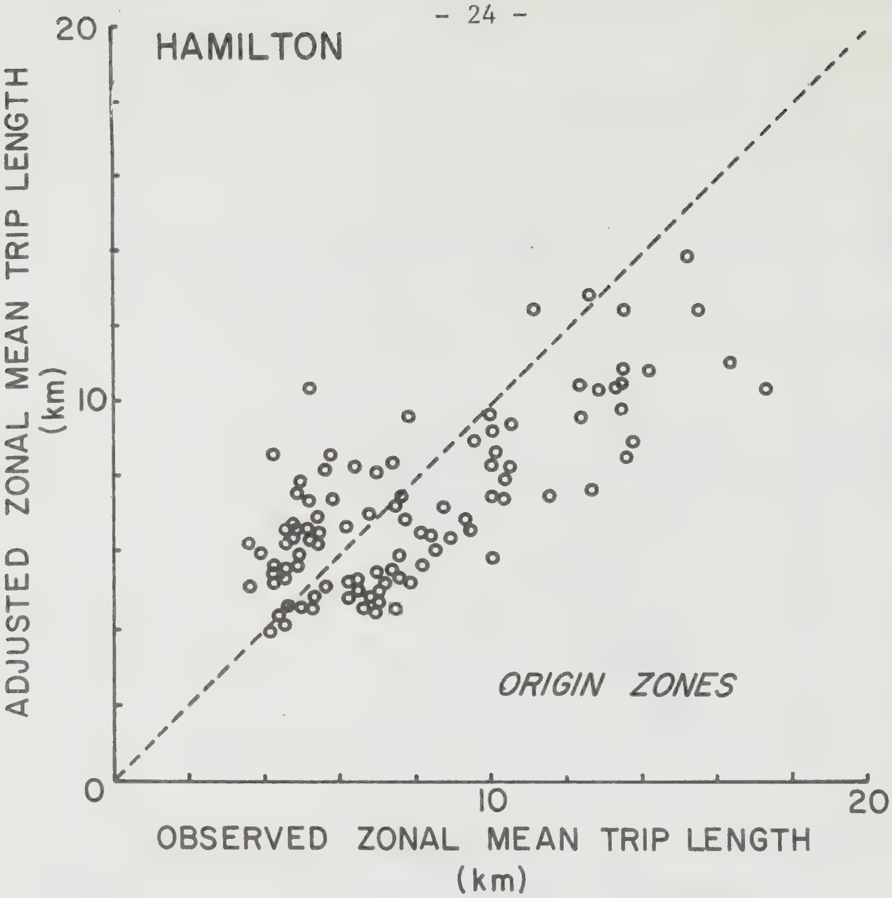


FIGURE 7. Adjusted vs Observed Origin Zone Mean Trip Lengths for Kitchener and London



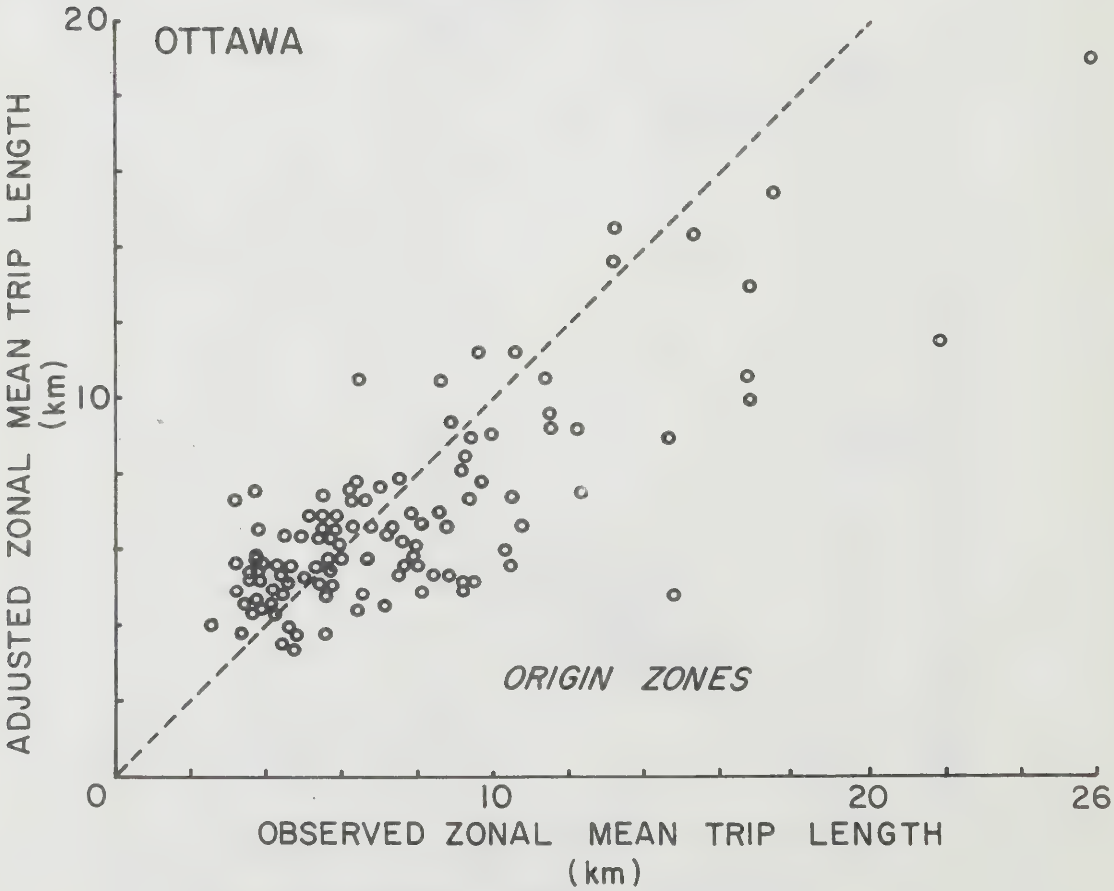
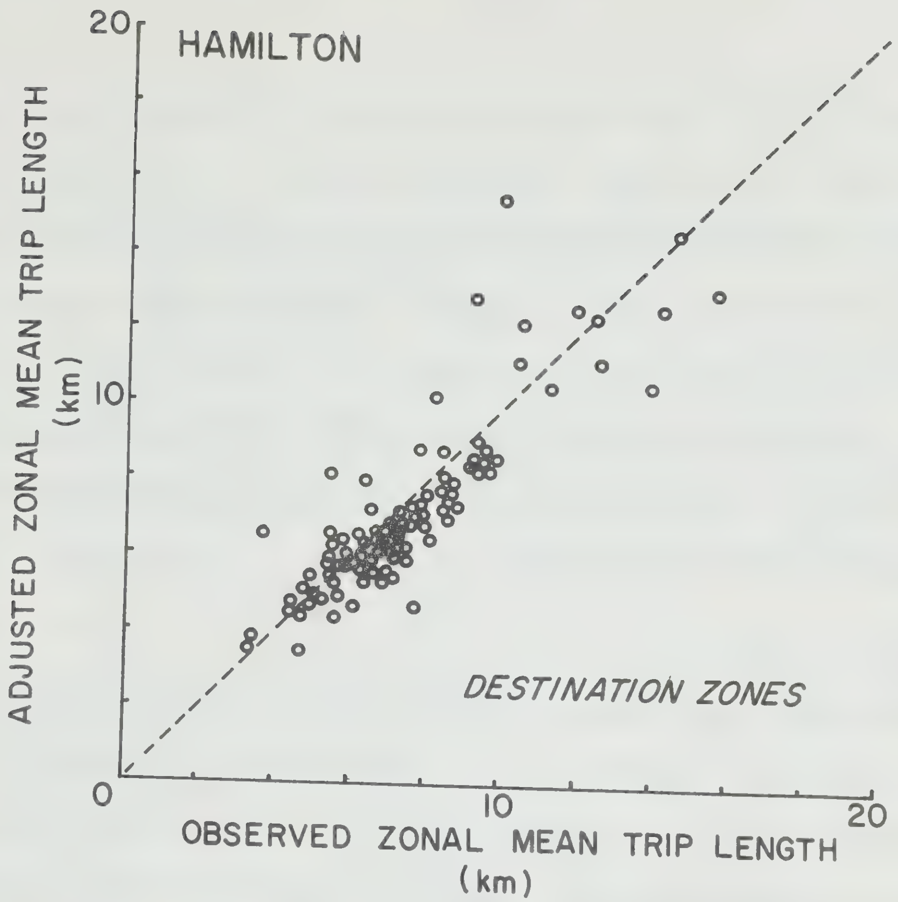


FIGURE 8. Adjusted vs Observed Origin Zone Mean Trip Lengths for Hamilton and Ottawa



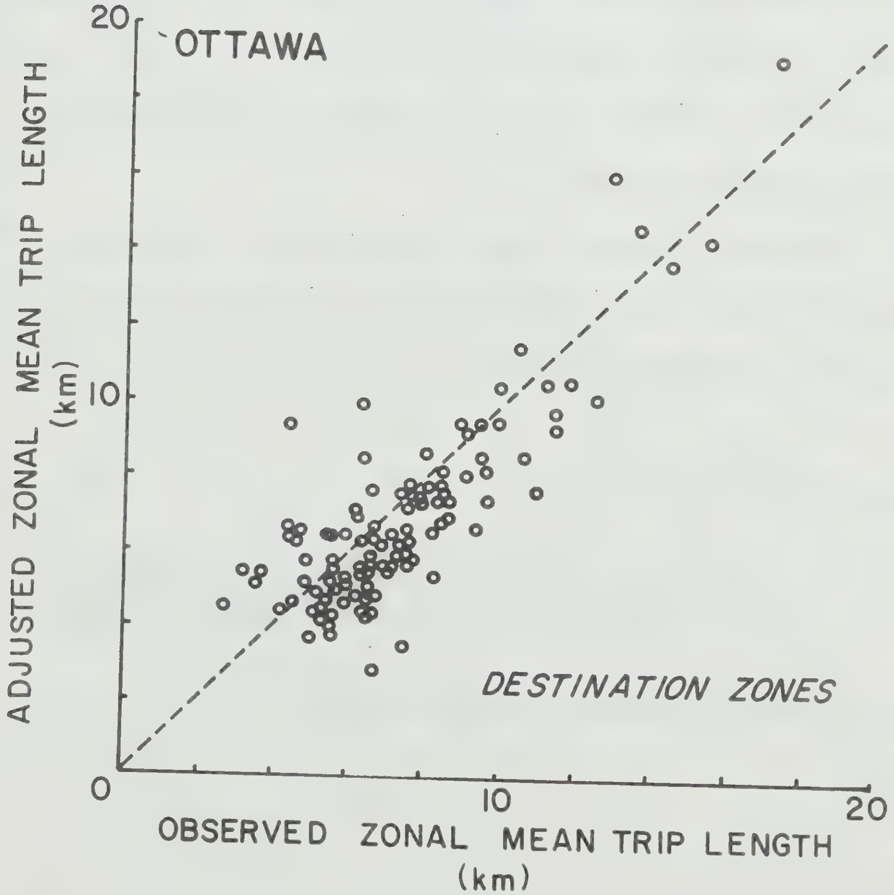


FIGURE 9. Adjusted vs Observed Destination Zone Mean Trip Lengths for Hamilton and Ottawa

Figures 7 to 8 illustrate the tendency for shorter trip lengths to become longer and the longer trip lengths shorter. The largest shifts in the mean trip lengths are associated with zones located on the periphery or the smaller-sized zones. Figure 9 illustrates that the mean trip lengths for the destination zones tend to be reduced by the balancing procedure. It should be remembered that the aim of the bi-proportional balancing procedure is not to preserve the trip length structure but the interaction structure of a community.

2.2 Clustering Techniques

The clustering technique used in this investigation is a procedure known as Ward's method which is a hierarchical agglomerative technique [4]. This technique begins with the n census tracts in a census area and groups them into a series of clusters with aggregation continuing until all of the census tracts are grouped into one cluster. With this technique the analyst is not constrained to some arbitrary pre-determined number of clusters.

The clustering technique uses a measure of the Euclidean distance between the destination-zone similarities of two census tracts which is defined in the following way:

$$d_{ij} = \left[\sum_{k=1}^{n} \frac{(T_{ik}^{*} - T_{jk}^{*})^2}{n}\right]^{1/2} \qquad (1)$$

where d_{ij} = the Euclidean distance between the destination characteristics of the two census tracts i and j

k = the destination zone under consideration for a particular origin zone (i or j)

T^*_{ik} = the normalized number of trips between origin zone i and destination zone k

n = the total number of census tracts

That is, d_{ij} is a measure of the similarity in the destination vectors of the two particular origin zones being considered for a potential merger. If the destination vectors were identical then $d_{ij} = 0$. The lower the magnitude of d_{ij} the greater the similarity between the two census tracts.

The clustering technique proceeds by first calculating an upper triangular matrix of the similarities in the destination vectors of each pair of census tracts. The pair of census tracts with the minimum d_{ij} magnitude are then merged into a cluster. The entries in the matrix of similarity are then re-calculated to reflect this merging of two census tracts and the procedure is continued until all census tracts are merged into one cluster. The logic of the clustering procedure is illustrated in Figure 10.

The calculations of the new entries in the similarity matrix do not involve a complete re-calculation of the entries in the similarity matrix since not all entries are affected. Suppose two census tracts i and j join to form a new group ij, usually labelled with the smallest census tract number. The similarity between this new cluster and another cluster, or census tract, is given by:

$$d_{h,ij} = \alpha_i d_{hi} + \alpha_j d_{hj} + \beta d_{ij} \qquad (2)$$

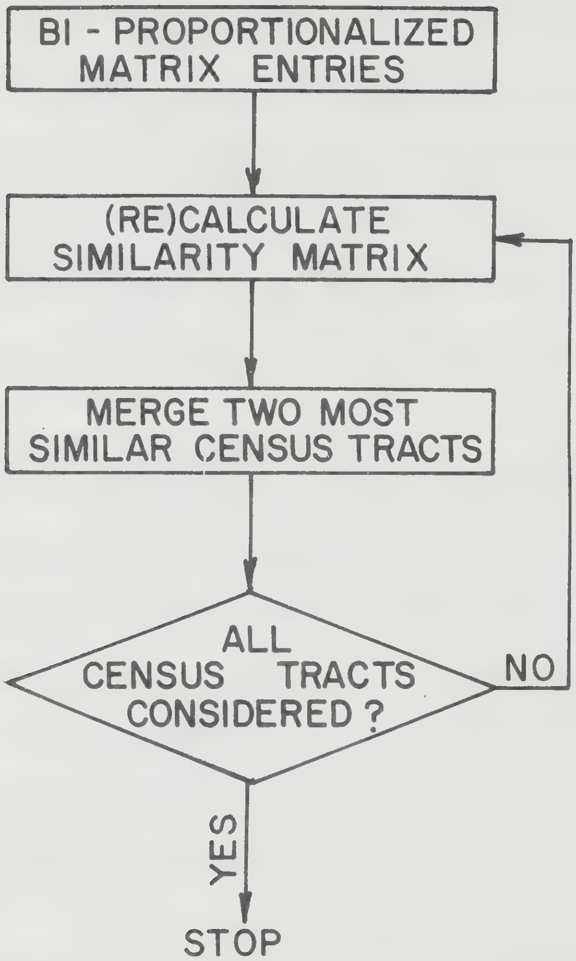


FIGURE 10. Steps in the Clustering Procedure

where $$\alpha_i = \frac{n_h + n_i}{n_h + n_i + n_j}$$

$$\alpha_j = \frac{n_h + n_j}{n_h + n_i + n_j}$$

$$\beta = \frac{-n_h}{n_h + n_i + n_j}$$

n = the number of census tracts in the cluster identified by the sub-script

Equation (2) indicates that the new measure of similarity is calculated from the weighted sum of the Euclidean distances between the pairs of clusters used in the previous step. The larger n_i is relative to n_j then the greater d_{hi} will be weighted in the calculation of the new similarity measure $d_{h,ij}$. Further, the larger the value of n_h then the smaller the weighting of d_{ij} which is the distance between the two census tracts being merged.

The distance value calculated by equation (1) may be interpreted as the within cluster sum of squared differences between the destination vectors of two census tracts. As more census tracts are included in a cluster then the distance value as calculated by equation (2) represents the within cluster sum of squared differences between a cluster centroid and the location of the individual census tracts in the cluster. The clustering procedure groups census tracts so as to minimize the increase in the sum of squared differences across the entire set of census tracts. One would expect that the first stages in the clustering hierarchy would consist of a number of pairwise census tract clusters each with a small increase in the sum of squared differences. Large increases in the aggre-

gate sum of squared differences might be expected as clusters are grouped to form eventually a single cluster.

2.2.1 An Example

Figure 11(a) shows a simple five zone home to work matrix of the type available from the census. This matrix may be transformed into the doubly-balanced matrix shown in Figure 11(b) by alternatively scaling the row and column entries so that they sum to 20, an arbitrary constant.

The entries in the matrix of Figure 11(b) may be used along with equation (1) to produce the upper triangular matrix of similarity measures shown in the upper part of Figure 12. For example:

$$d_{12} = \left[\frac{(8-3)^2 + (7-8)^2 + (1-4)^2 + (4-4)^2 + (0-0)^2}{5}\right]^{1/2} = 2.645$$

and

$$d_{13} = \left[\frac{(8-4)^2 + (7-3)^2 + (1-7)^2 + (4-1)^2 + (0-7)^2}{5}\right]^{1/2} = 5.020$$

and so on.

Inspection of the initial similarity matrix of Figure 12 shows that the minimum similarity coefficient is d_{35} with a magnitude of 2.49. The first step in the clustering procedure, then, is to group zones 3 and 5 to form a new zone 3' , where the marginal increase in the error be updated to reflect this initial clustering of zones 3 and 5 and this requires the calculation of $d_{13'}$, $d_{23'}$ and $d_{34'}$. Equation (2) may be used to calculate these new similarities:

$$d_{13'} = \alpha_3 d_{13} + \alpha_5 d_{15} + \beta d_{35}$$

$$\alpha_3 = \frac{n_1 + n_3}{n_1 + n_3 + n_5} = \frac{2}{3}$$

		WORK ZONE					
		1	2	3	4	5	
HOME ZONE	1	14	3	1	5	0	23
	2	5	4	3	5	0	17
	3	10	2	5	1	2	20
	4	5	0	4	15	1	25
	5	6	1	2	4	2	15
		40	10	15	30	5	

		WORK ZONE					
		1	2	3	4	5	
HOME ZONE	1	8	7	1	4	0	20
	2	3	8	5	4	0	20
	3	4	3	6	1	6	20
	4	2	0	5	8	5	20
	5	3	2	3	3	9	20
		20	20	20	20	20	

FIGURE 11. Doubly Balanced Home to Work Linkages Matrix

INITIAL SIMILARITY MATRIX

HOME ZONE	WORK ZONE 1	2	3	4	5
1		2·65	5·02	5·02	5·55
2			4·31	4·27	5·25
3				4·05	2·49
4					4·07

REVISED SIMILARITY MATRIX

HOME ZONE	WORK ZONE 1	2	3'	4	5
1		2·65	6·22	5·02	
2			5·54	4·27	
3'				4·57	
4					

REVISED SIMILARITY MATRIX

HOME ZONE	WORK ZONE 1	2	3'	4	5
1'			7·5	5·3	
2					
3'				4·5	
4					

FIGURE 12. Successive Similarity Matrices

$$\alpha_5 = \frac{n_1 + n_5}{n_1 + n_3 + n_5} = \frac{2}{3}$$

$$\beta = \frac{- n_1}{n_1 + n_3 + n_5} = \frac{1}{3}$$

therefore

$$d_{13'} = \frac{2}{3} \cdot 5.02 + \frac{2}{3} \cdot 5.55 - \frac{1}{3} \cdot 2.49 = 6.22$$

$$d_{23'} = 5.54$$

$$d_{3'4} = 4.57$$

The revised similarity matrix which reflects the merging of zones 3 and 5 is shown in Figure 12. It should be re-called that $d_{13'}$ represents the similarity between zone 1 and the new zone 3' formed from the clustering of zones 3 and 5.

Inspection of the revised similarity matrix shows the minimum distance is $d_{12} = 2.65$ and the second step in the clustering procedure is to merge zones 1 and 2 to form a new zone 1' with a marginal increase in error sum of squares of 2.65. A new similarity matrix has to be produced from the calculation of $d_{1'3'}$ and $d_{1'4}$ and this matrix is shown in the lower part of Figure 12. The minimum distance measure in this revised matrix is $d_{3'4}$ which means that the third clustering step involves merging 3' and 4 to form a new zone 3" with a marginal increase in the error sum of squares of 4.5. Once again a new similarity matrix has to be calculated which has one entry of $d_{1'3''} = 7.38$. The final fusion into one cluster occurs with a marginal increase in the

error sum of squares of 7.38. The clustering sequence for this example is zones 3 and 5, then zones 1 and 2, then merged 3 and 5 with 4 and finally merged 1 and 2 with 3, 4 and 5.

Figure 13 shows the dendrogram for this particular example where this dendogram has been plotted using the convention of the computer program used in this study. The clusters are arranged as far as possible in terms of increasing zone number magnitude. The horizontal lines representing the fusion of two census tracts, or clusters, is drawn at the magnitude of the marginal increase in the sum of squares at which this union occurs where the magnitude is shown as the ordinate. That is, zones 3 and 5 are clustered first at a magnitude of 2.49; zones 1 and 2 are then clustered at a magnitude of 2.65, and so on. The lower part of Figure 13 traces the marginal increase in the error sum of squares as the number of clusters decreases to 1. This diagram is simply a restatement of the vertical spacings of the horizontal lines shown in the upper part of the diagram.

2.3 Dendrograms for Ontario Census Areas

The appendix to this chapter shows the dendrogram constructed for each of the fifteen Ontario census areas. These dendrograms have been plotted using the convention introduced in Figure 13. The dendrograms trace the clustering hierarchy and identify the magnitudes of the error sums of squares at which merging occurs. The dendrograms for the Guelph and Brantford census areas are illustrated in Figure 14 while the increases in the error sums of squares with decreasing numbers of clusters are shown in Figure 15.

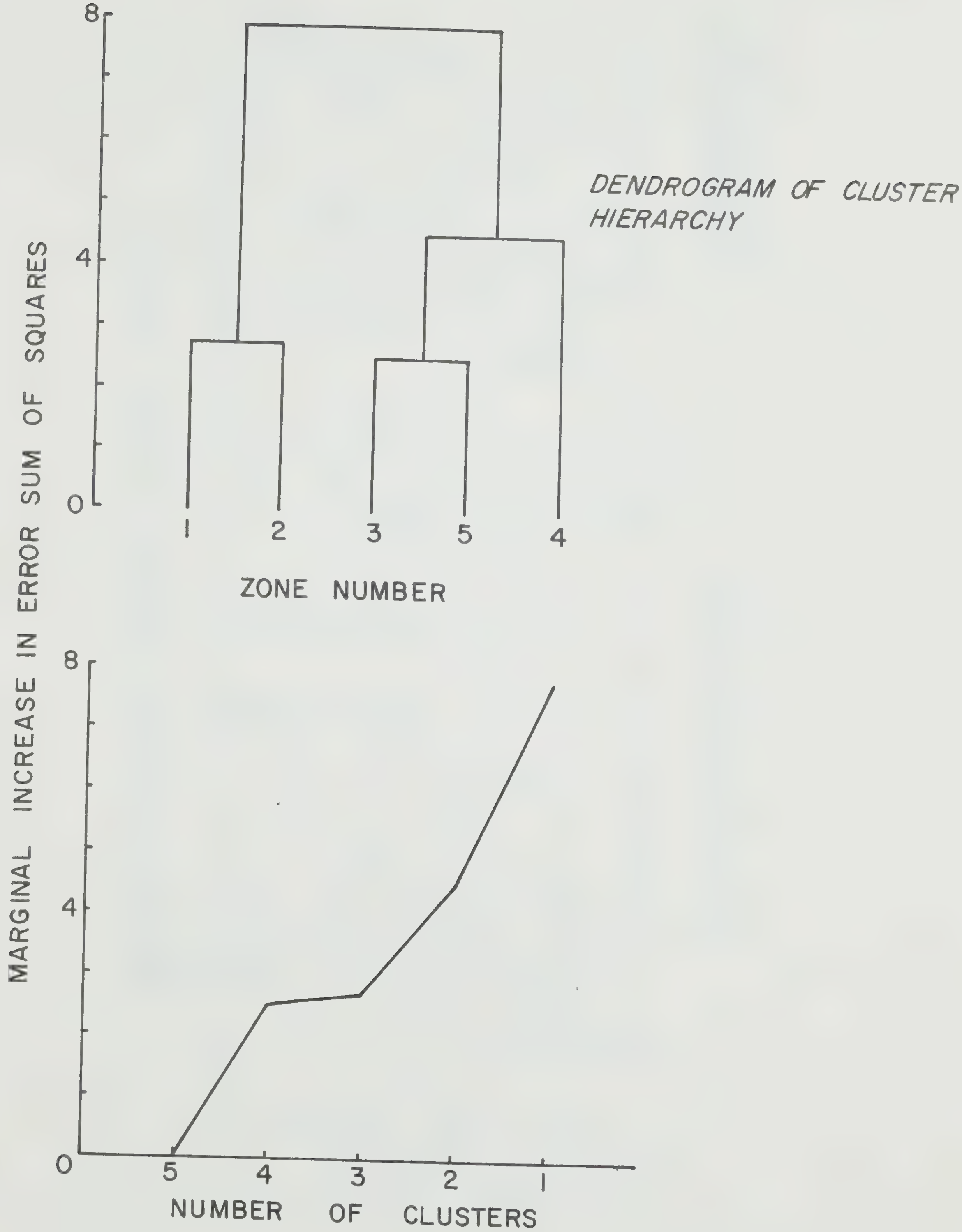


FIGURE 13. Dendrogram of Cluster Hierarchy and Associated Marginal Increase in Error Sum of Squares

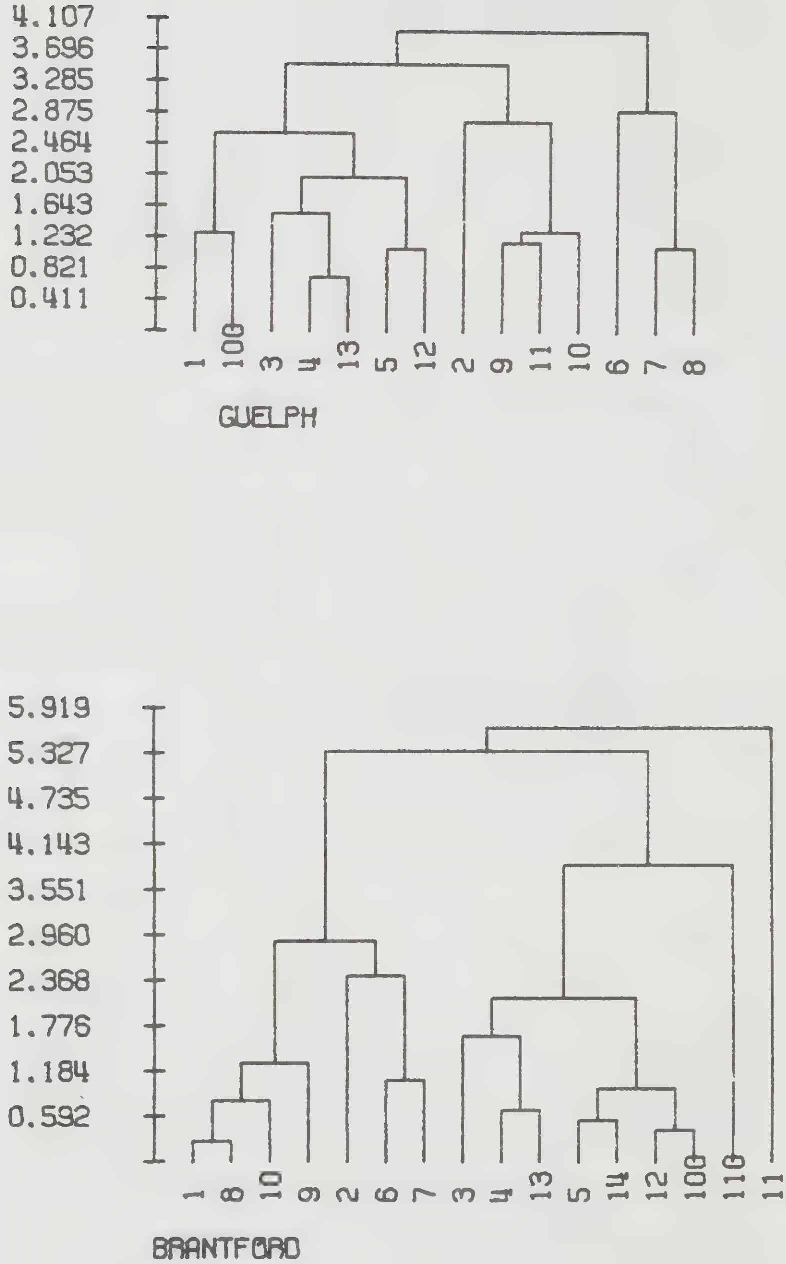


FIGURE 14. Dendrograms for Guelph and Brantford

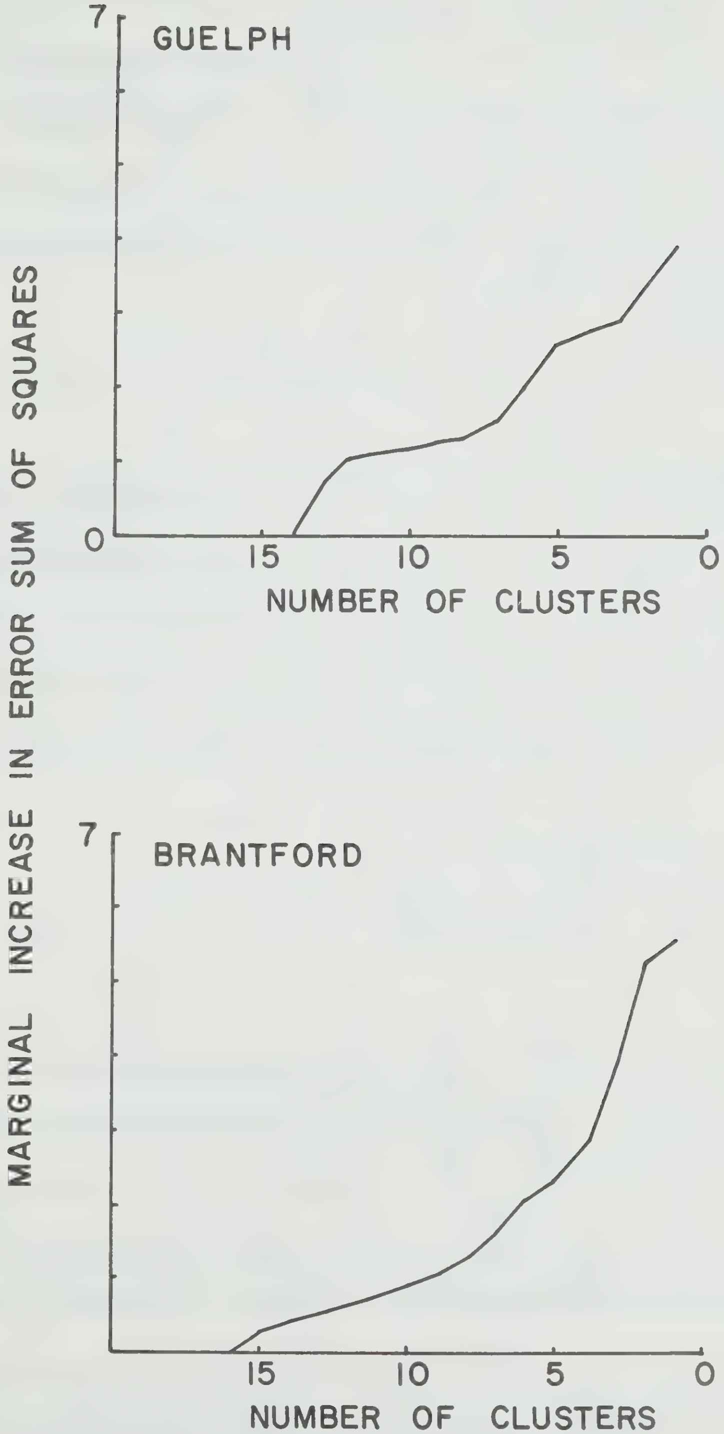


FIGURE 15. Marginal Increase in Error Sum of Squares with Decreasing Numbers of Census Tracts for Guelph and Brantford

Inspection of the dendrogram for Guelph in Figure 14 shows that the first pair of census tracts to cluster are 4 and 13, followed by 7 and 8, 5 and 12, and 9 and 11. As the clustering process continues 10 clusters with the already clustered 9 and 11. This dendrogram shows that clustering continues without major jumps in the error sum of squares. With three clusters the census tract composition of each cluster is:

I : 1, 100, 3, 4, 13, 5 and 12

II : 2, 9, 11 and 10

III : 6, 7 and 8

The clustering process in the Brantford region has slightly different characteristics than for Guelph. While the initial stages of the clustering are similar the error sum of squares increases sharply during the final stages of the clustering process. The dendrogram for Brantford suggests the following census tract composition for three clusters:

I : 1, 8, 10, 9, 2, 6 and 7

II : 3, 4, 13, 5, 14, 12, 100 and 110

III : 11

It is interesting to note that census tract 11 in Brantford is almost exclusively an industrial zone and that it clusters with the remaining census tracts only at the last stage. It should also be noted from Figures 14 and 15 that the increase in the error sum of squares is large when clusters I and II are merged and that a similarly large increase occurs when census tract 110 is merged with the other members of cluster II in the previous stage. Census tract 110 embraces the the Paris area which is part of the Brantford census area and its

forced merger with the census tracts in the northwestern part of Brantford produces the relatively large increase in the error sum of squares at the anti-penultimate stage of clustering.

The characteristics of the Guelph and Brantford dendrograms are sufficient to illustrate the general properties of the dendrograms contained in the appendix of this chapter. Some of the effects highlighted above are more extreme in the larger census areas than in the two smaller census areas. For example, the Kitchener census area consists of the three major municipalities of Waterloo, Kitchener and Cambridge which embrace relatively self-contained commuter sheds. When these areas are forced together large increases in the error sums of squares occur. The ordinates of all of the dendrograms presented in the appendix are plotted to the same scale and differences in the clustering properties of each census area are detected easily. The residential zone clustering properties of each of the fifteen Ontario are discussed in detail in the following paragraphs. It should be recalled that the residential zones have been clustered on the basis of similarities in destination vectors and that no census tract adjacency constraints have been used.

In reviewing the commuting structure of each census area five broad categories of effects have been used to assist in this analysis and these categories are:

Multi-Community Composition: several census areas embrace a number of separate municipalities where there is some degree of commuting self-containment.

Topographic and Man-Made Influences: many census areas are bisected by rivers and other topographic features as well as certain man-made features such as rail lines which may influence commuting patterns.

Timing of Development: strong linkages may be observed between residential and employment zones that develop at about the same time even though these zones may be separated by some distance.

Socio-Economic Factors: certain dwelling unit types and employment sectors may attract higher than average proportions of particular socio-economic groups and their specific locations will condition spatial linkage patterns.

Specific Employment Concentrations: several census areas have very strong concentrations of employment of a specific type in one or two census tracts and specific commuting patterns may be associated with these areas.

Guelph

Figure 16 shows the census tract grouping in Guelph for three clusters. Cluster I consists of the six census tracts along the eastern side of the area, cluster II consists of the four census tracts on the western side and cluster III embraces the CBD (CT 6) and the two census tracts immediately to the west of the CBD. The peripheral clusters I and II are linked with job opportunities in CT 1 (University of Guelph) and CT 9 (suburban industrial area) while the census tracts in cluster III are linked primarily to jobs in the CBD. The census tracts marked with an * in Figure 16 experienced significant population growth during 1966-1971 and developed strong linkages with employment in CTs 1 and 9. The primary factor influencing spatial linkages in Guelph is the timing of development with the inner suburbs oriented to CBD employment and the newer outer suburbs linked with suburban employment. Figure 17 shows the major home to work linkages observed in Guelph in 1971 which also illustrates these two broad commuting pattern differences.

Peterborough

Figure 18 shows the census tract grouping in Peterborough for three clusters. CT 14 had a very small population in 1971 and it does not cluster naturally with the other Peterborough census tracts. The dendrogram in the appendix shows that it clusters with CT 110 which includes the Lakefield Community north of Peterborough. Peterborough had two major concentrations of employment in 1971 and these were the CBD in CT 7 and the manufacturing employment in CT 4. Cluster I includes the older peripheral residential zones adjacent to the manufacturing plants in CT 4 along with CTs 11 and 12 which experienced sigificant population growth during 1966-1971. The principal determinants of this cluster are strong linkages with jobs in CT 4. In contrast the CTs in Cluster II have strong linkages with jobs in CT 7. The CTs in this cluster include the older established residential areas in CTs 5, 6, 7, 8 and 10 as well as the newer residential areas in CTs 9 and 13. The principal determinant of spatial linkages in Peterborough seems to be socio-economic with some influences of the timing of development. Figure 19 shows the major home to work linkages in Peterborough.

Sarnia

Figure 20 shows the three major clusters of census tracts in Sarnia. Cluster I consists only of CT 1 which contains the concentration of petro-chemical jobs in Sarnia and only a small number of households. Cluster II contains CTs 2,3,5, 6, 7 (CBD) and 100 (railway jobs). These CTs tend to have strong linkages with jobs in CTs 7 and 100. Cluster III embraces the remaining CTs along the eastern boundary of Sarnia which cluster on the basis of their strong

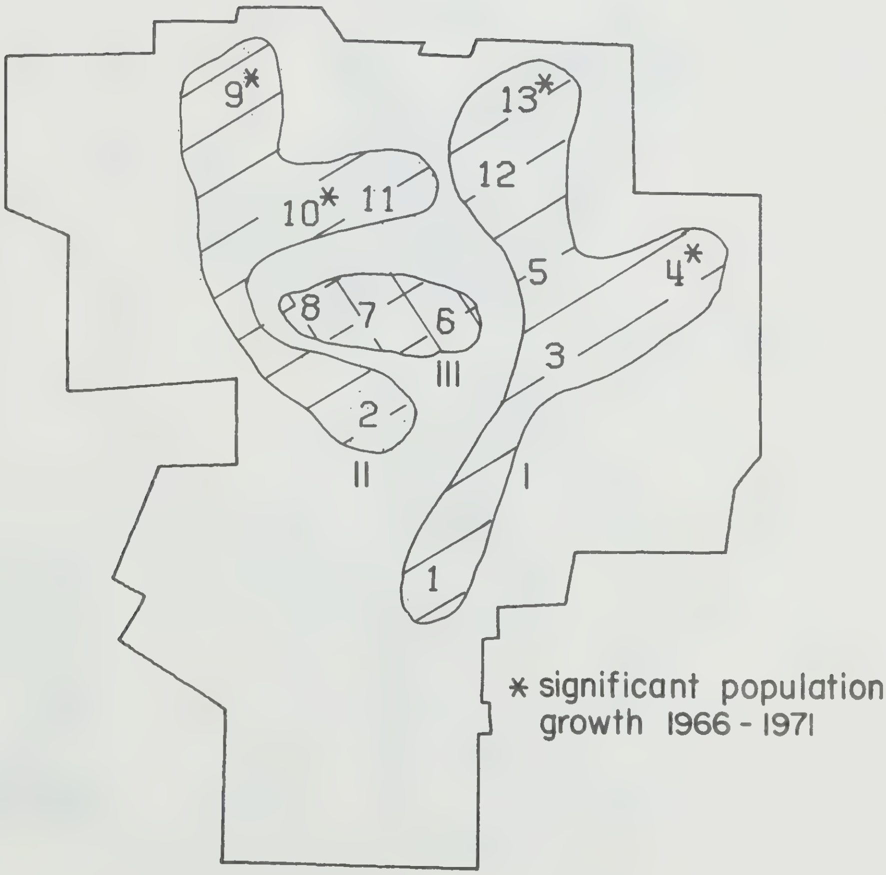


FIGURE 16. Census Tract Clusters for Guelph

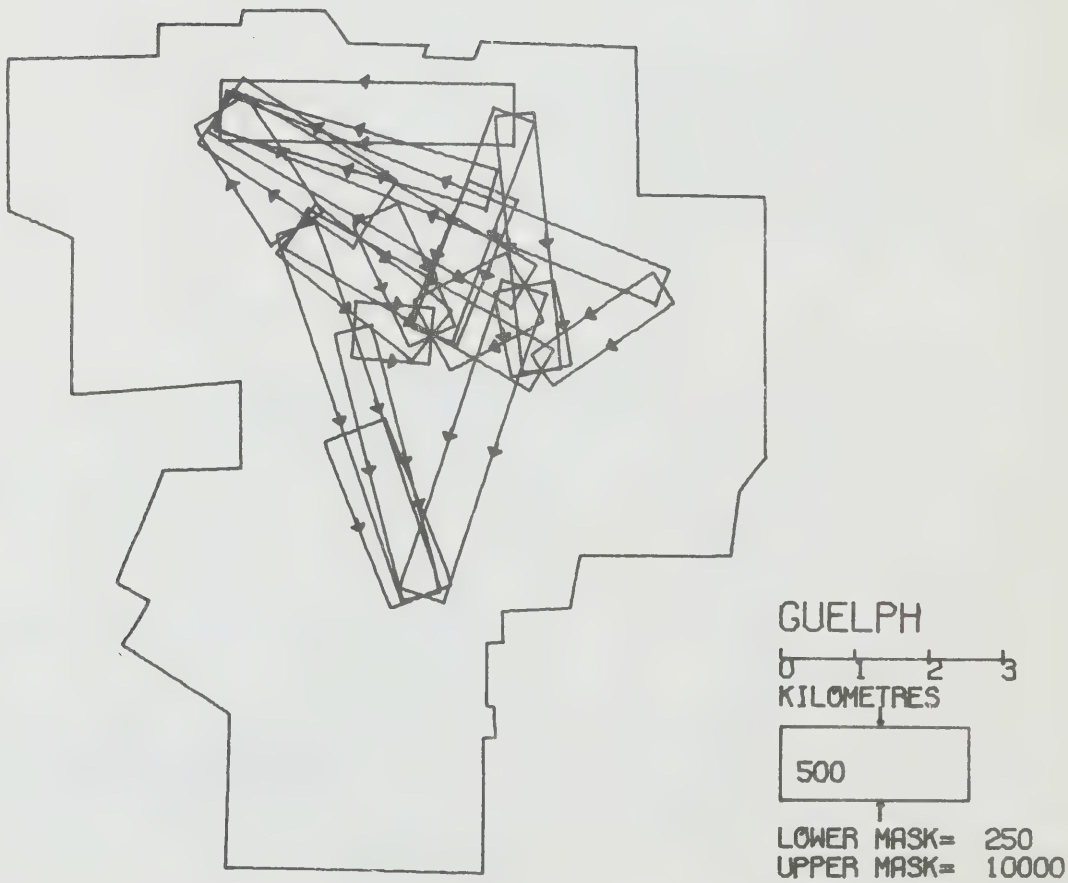


FIGURE 17. Major Home to Work Linkages for Guelph

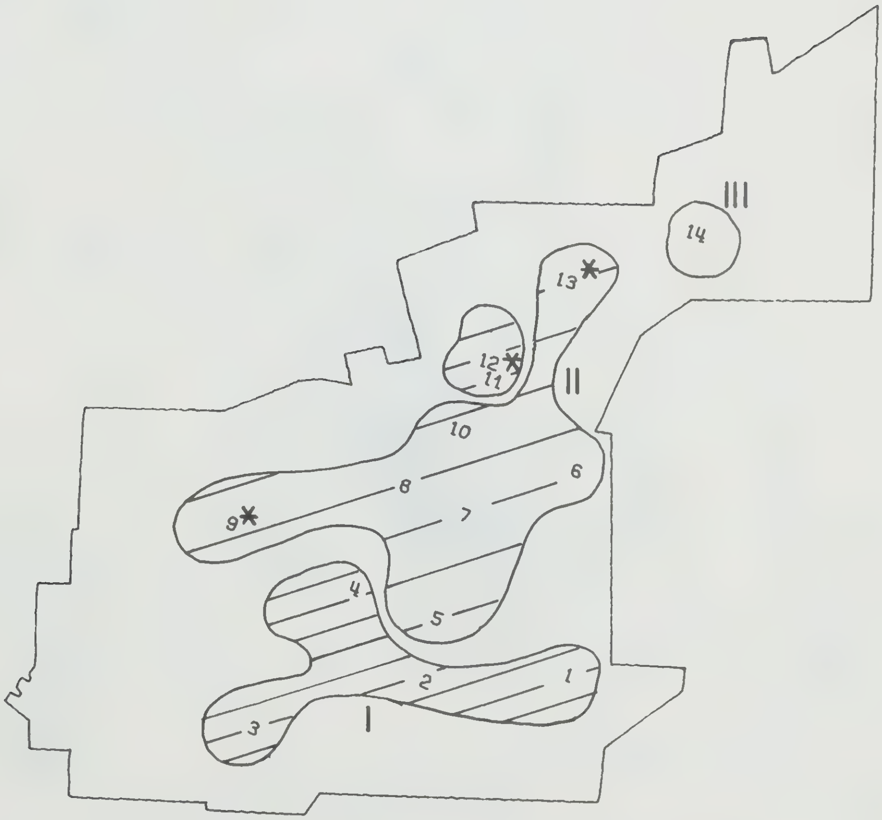


FIGURE 18. Census Tract Clusters for Peterborough

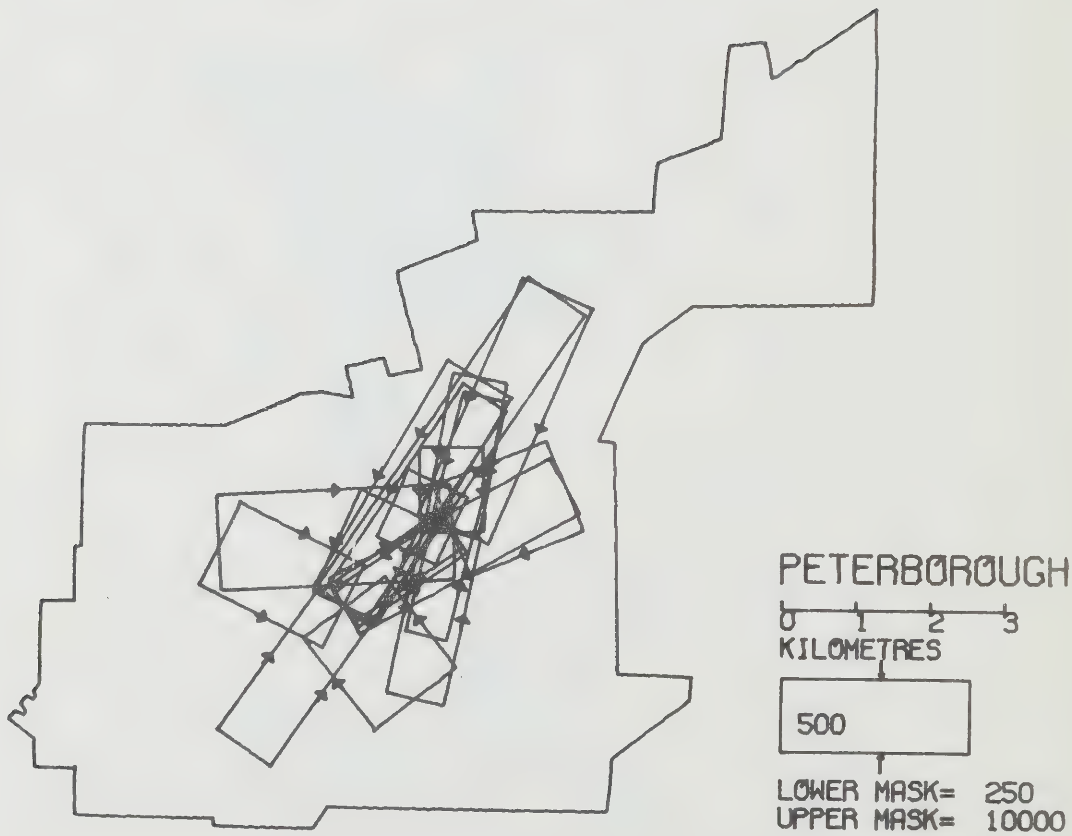


FIGURE 19. Major Home to Work Linkages for Peterborough

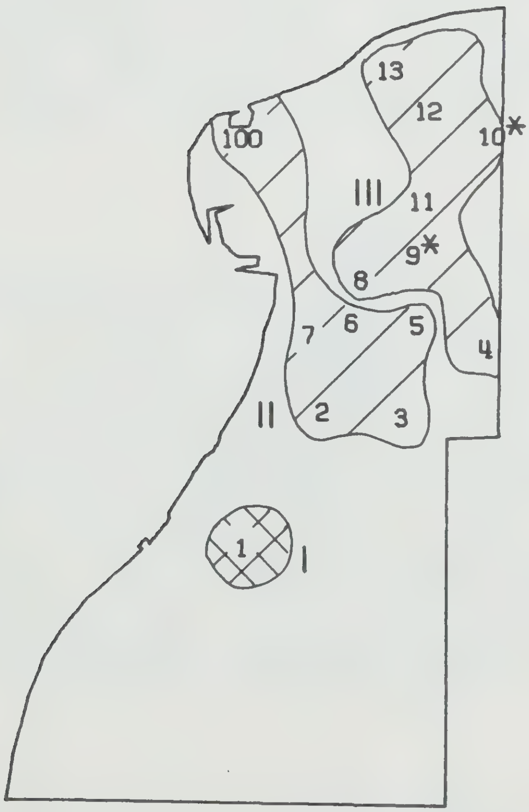


FIGURE 20. Census Tract Clusters for Sarnia

linkages with jobs in CT 1. Some separation of the CTs in cluster III may be detected from the Sarnia dendrogram in the appendix where differences are due primarily to linkages with jobs in CTs 7 and 100. The principal determinants of spatial linkages in Sarnia are socio-economic factors along with the strong concentration of jobs in the petro-chemical industry. Figure 21 illustrates the major trip linkages in Sarnia and these highlight the importance of the three major employment concentrations.

Brantford

Figure 22 illustrates the census tract composition of three clusters for Brantford where cluster III contains only CT 11 a major manufacturing zone with few households. Figure 22 shows that Brantford divides into two fairly distinct commuting clusters. While the significant amounts of employment in the central area CTs 5 and 6 are common to both commuter sheds the differentiation between the commuter sheds is created by the linkages to jobs in CTs 1 and 2 for cluster I and linkages to jobs in CTs 3 and 11 for the CTs in cluster II. The dendrogram in the appendix illustrates that there is a significant increase in the error sum of squares when clusters I and II join. The dendrogram shows that this large increase in error sum of squares is due to CT 110 (Paris) joining cluster II just prior to its union with cluster I. The principal determinants of the clusters in Brantford are its multi-community composition and the presence of the Grand River which provides part of the boundary between clusters I and II. Figure 23 illustrates the principal home to work linkages in Brantford.

Sault Ste. Marie

Figure 24 shows two major clusters of census tracts for Sault Ste. Marie. Employment in 1971 was concentrated in CT 8 (steel mill complex) and CT 6 (CBD). Because of the magnitude of the jobs concentrated in CT 8 all residential zones have relatively strong linkages with this CT. The CTs in cluster I are in the eastern part of the area and they have strong linkages with the CBD employment as well as the steel mill complex. CTs in cluster II have very strong linkages to the steel mill complex. The dendogram in the appendix shows that there is more heterogeneity in the destination characteristics of CTs in cluster II than in cluster I and that there is a sharp increase in the error sum of squares when clusters I and II are merged. The principal determinant of commuting patterns in Sault Ste. Marie is the large concentration of jobs in the steel mill complex with socio-economic factors influencing some of the linkage patterns. Figure 25 illustrates the major trip linkages in Sault Ste. Marie and shows the dominance of the steel mill oriented trip linkages.

Kingston

Figure 26 shows three major clusters of census tracts for the Kingston area. CT 13 is essentially a manufacturing employment zone which only clusters with the other CTs in the final stage creating a sharp increase in the error sum of squares. Cluster I contains the CTs in the older inner suburbs of Kingston which have strong linkages with Queen's University (CT 2), the hospitals (CTs 2 and 1) and the CBD (CT 1). Clusters II and III embrace CTs that are located

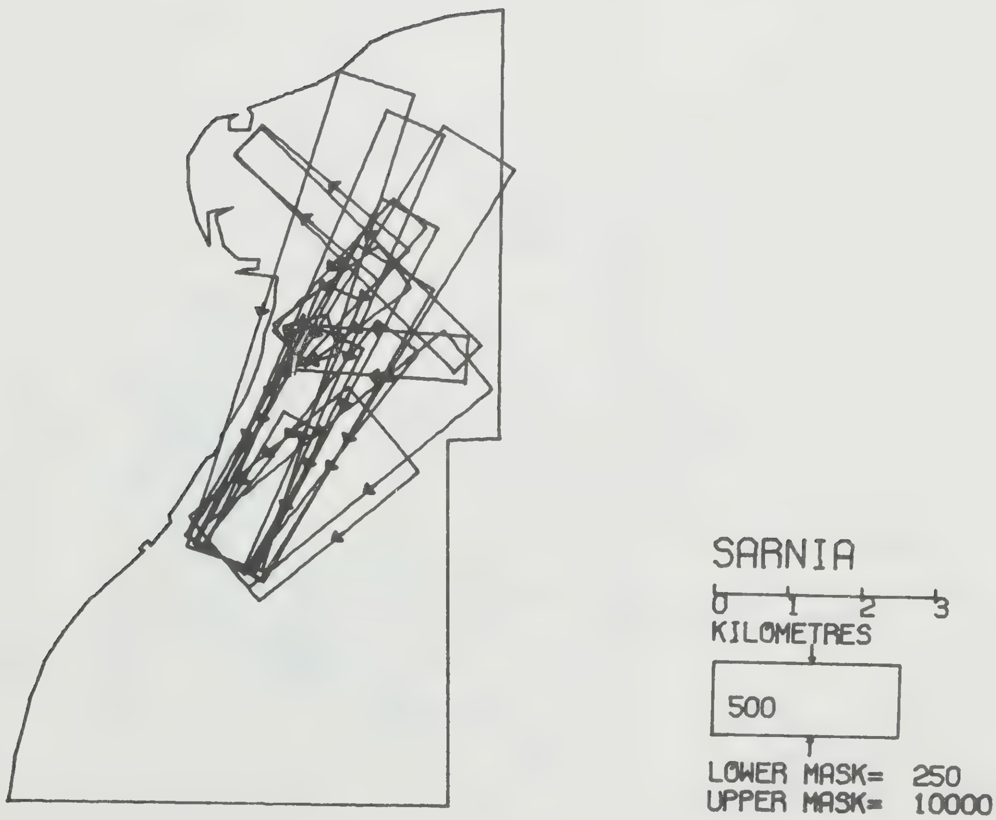


FIGURE 21. Major Home to Work Linkages for Sarnia

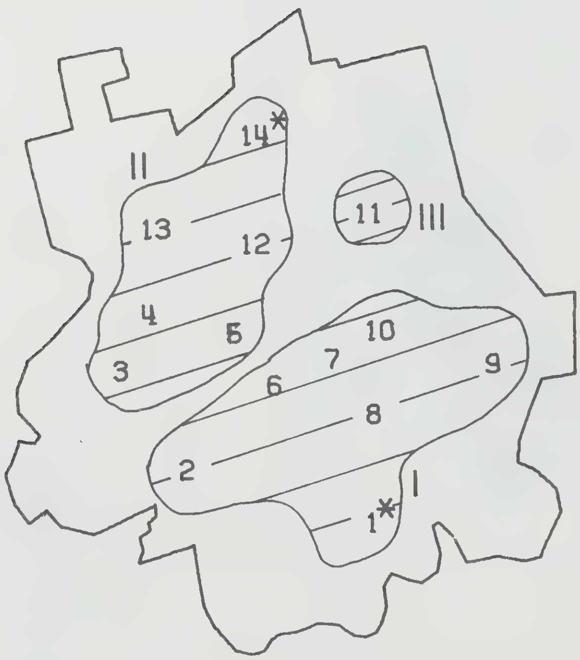


FIGURE 22. Census Tract Clusters for Brantford

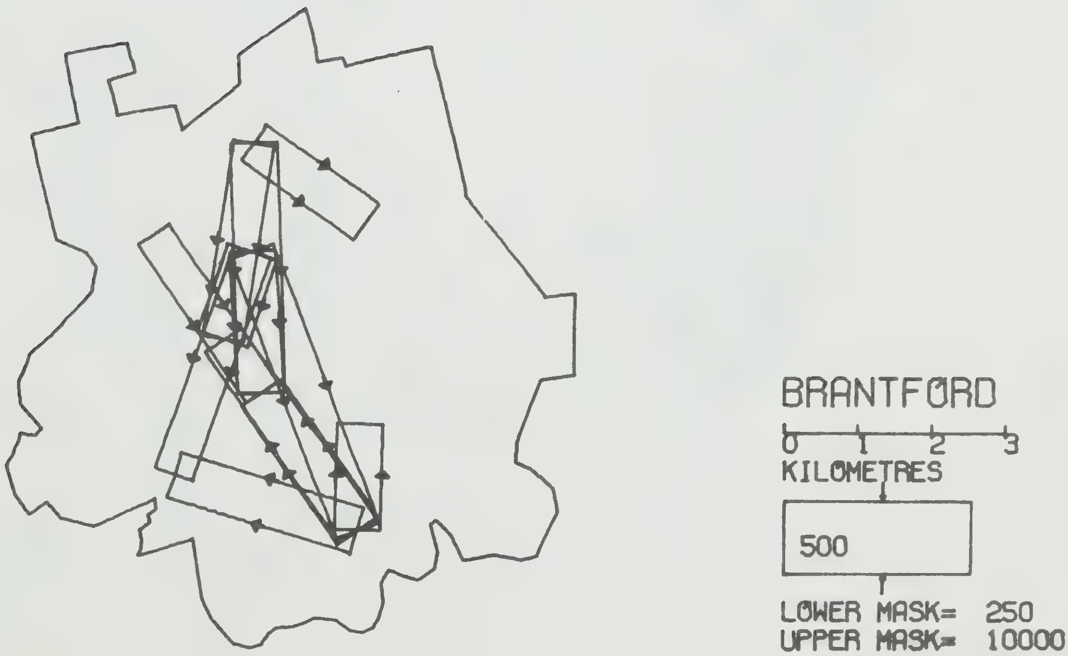


FIGURE 23. Major Home to Work Linkages for Brantford

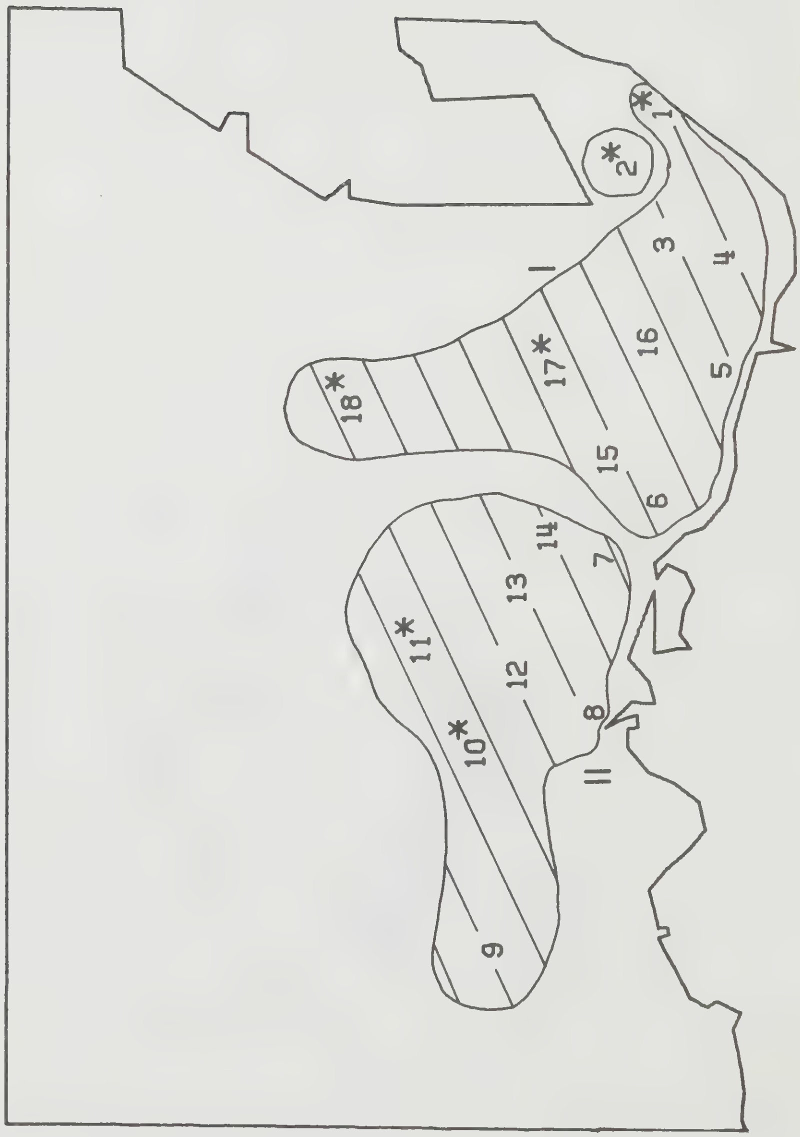


FIGURE 24. Census Tract Clusters for Sault Ste. Marie

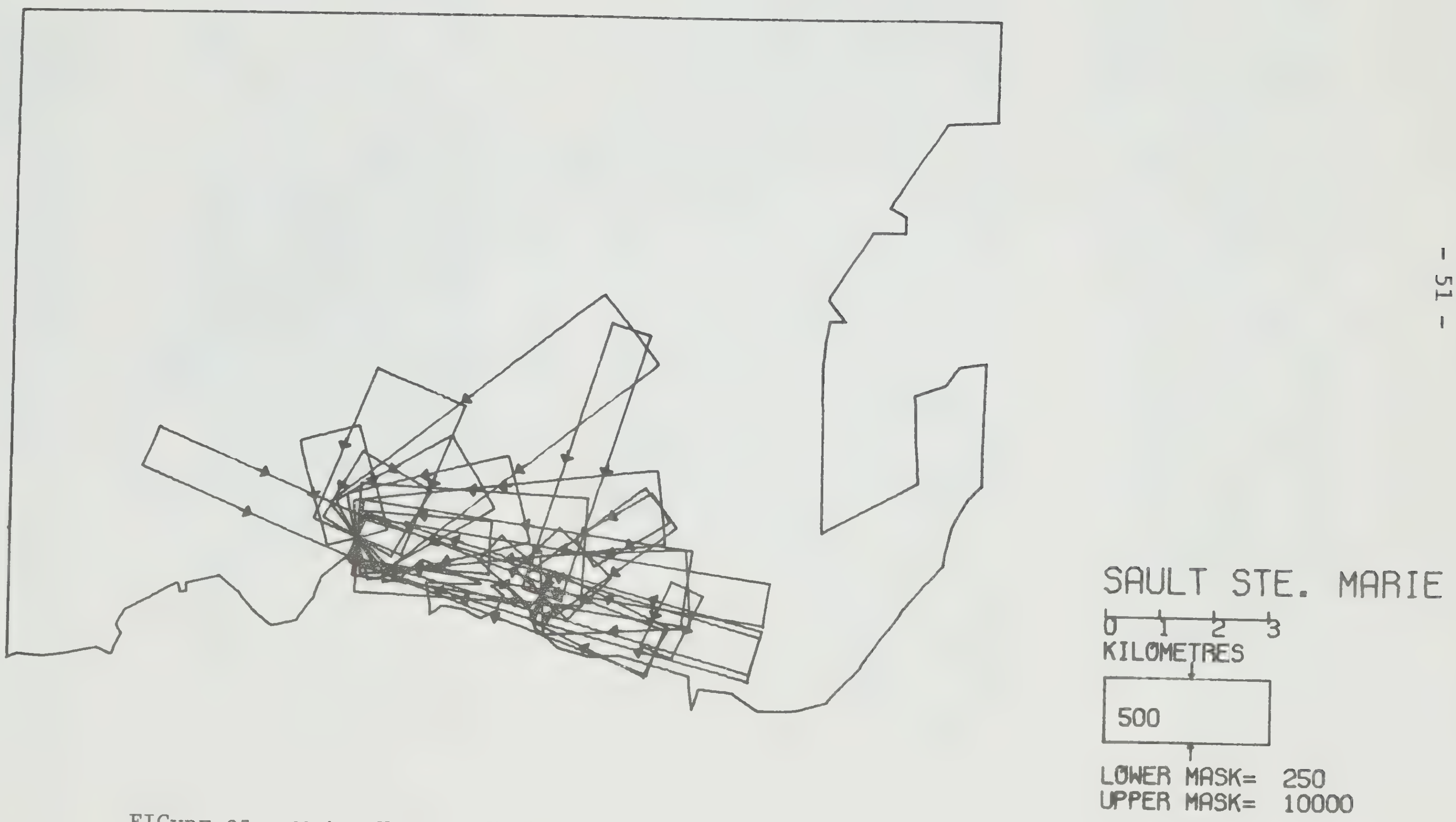


FIGURE 25. Major Home to Work Linkages for Sault Ste. Marie

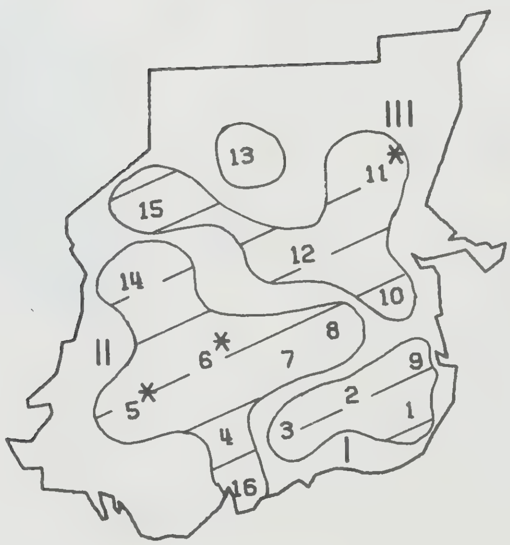


FIGURE 26. Census Tract Clusters for Kingston

to the north and south of Princess Street. While the CTs in these two clusters focus on different employment centres there are strong socio-economic differences between the two clusters of census tracts. Socio-economic factors are the most important determinant of home to work linkages in Kingston. Topographic constraints also influence commuting linkages from the residential areas in Kingston Township but these are not shown. Figure 27 shows the major home to work linkages in Kingston and illustrates the dominance of employment at Queen's University/Kingston General Hospital and the CBD.

Thunder Bay

Figure 28 shows the two major clusters of census tracts for Thunder Bay where these clusters embrace CTs within the former Cities of Fort William and Port Arthur, respectively. CTs 20 and 23 and outlying CTs 110 and 120 merge with the Fort William CTs in cluster I only after a significant increase in the error sum of squares. CT 20 is essentially an employment area and the residences in this CT exhibit an unusually long mean trip length. The dendrogram illustrates that destination patterns within each cluster are very homogeneous and that a very large increase in the error sum of squares occurs when clusters I and II are joined. Clearly, the major determinant of commuting linkages in the Thunder Bay area in 1971 were the former municipal boundaries. Within the former Fort William area employment in the CBD (CT 6) and CT 3 are the principal determinants of commuting patterns. The self-contained nature of the commuting patterns is further illustrated in Figure 29 which illustrates the home to work linkage patterns.

Oshawa

Figure 30 shows the three major census tract clusters for Oshawa where the boundaries of clusters I and II are essentially the municipal boundaries of Oshawa and Whitby, respectively. The dendrogram shows that cluster III (CTs 105, 110) merges with cluster II and that there is a sharp increase in the error sum of squares when clusters I and II join. Commuting patterns within Oshawa are dominated by the manufacturing employment concentrations. Municipal boundaries are the principal determinant of commuting patterns within the Oshawa census area and socio-economic factors tend to determine the spatial linkages patterns within Oshawa. In addition, the growth of housing opportunities in CT 105 north of Highway 401 has influenced the spatial linkages. Figure 31 illustrates the major home to work linkages in Oshawa.

Sudbury

Figure 32 shows the Sudbury census tracts grouped into two major clusters. The CTs in cluster I located in the southwest sector of the area have strong commuting linkages with the CBD. The CTs in cluster II are located in the northeast and tend to have strong linkages with the mining jobs. The dendrogram in the appendix identifies a third cluster of residential areas which are located in the townships surrounding Sudbury. A significant increase in the error sum of squares occurs when clusters I and II are clustered with

FIGURE 27. Major Home to Work Linkages for Kingston

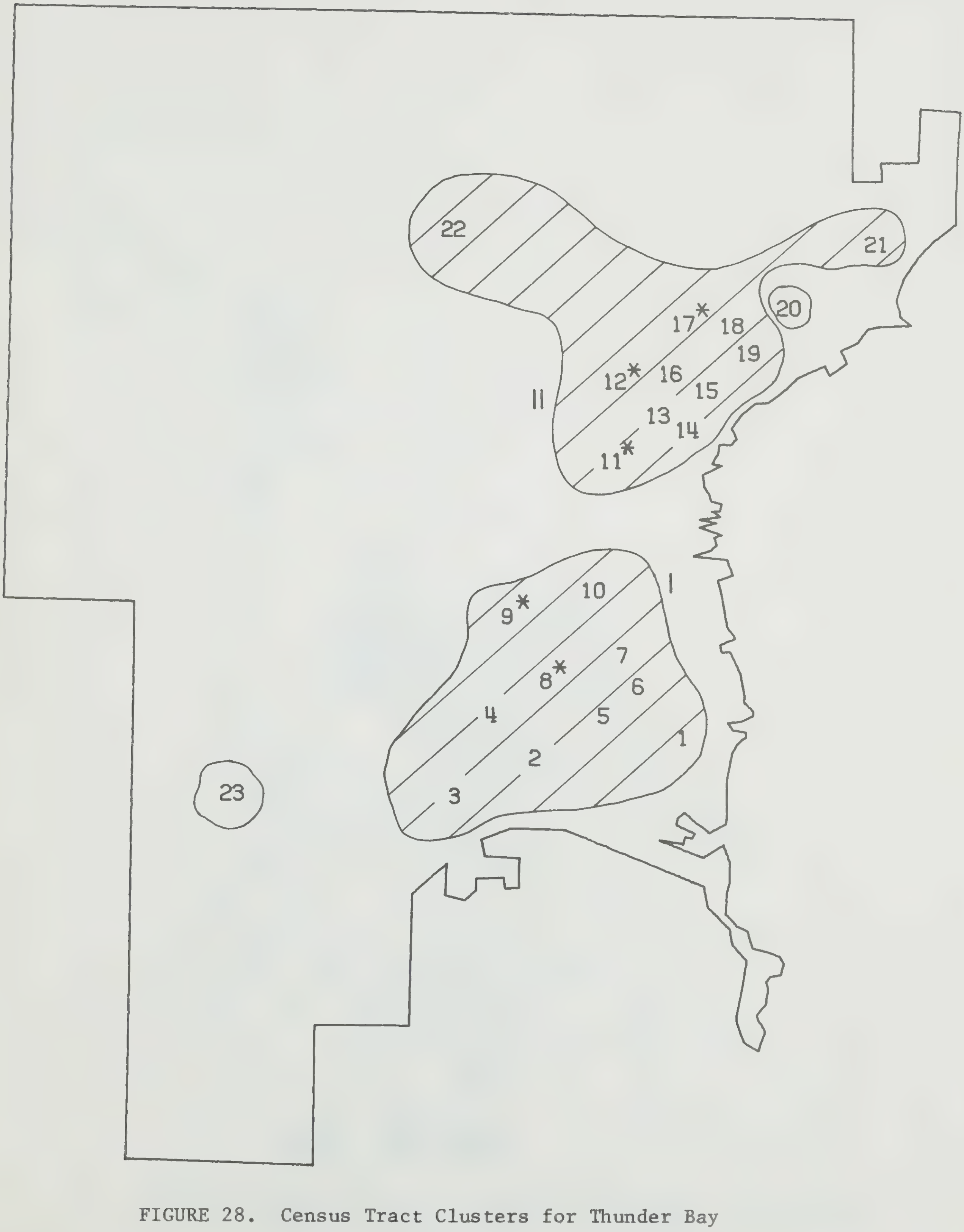


FIGURE 28. Census Tract Clusters for Thunder Bay

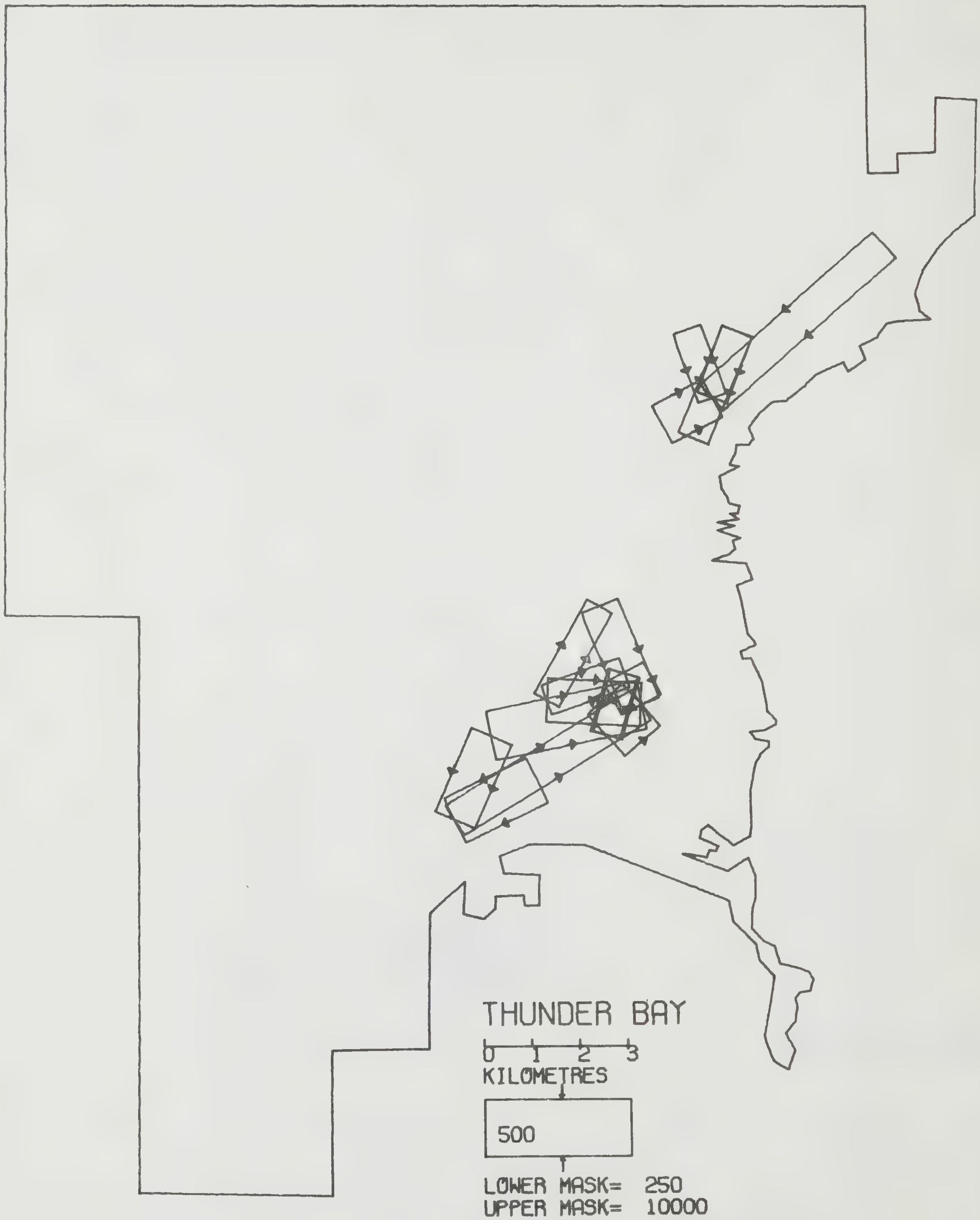


FIGURE 29. Major Home to Work Linkages for Thunder Bay

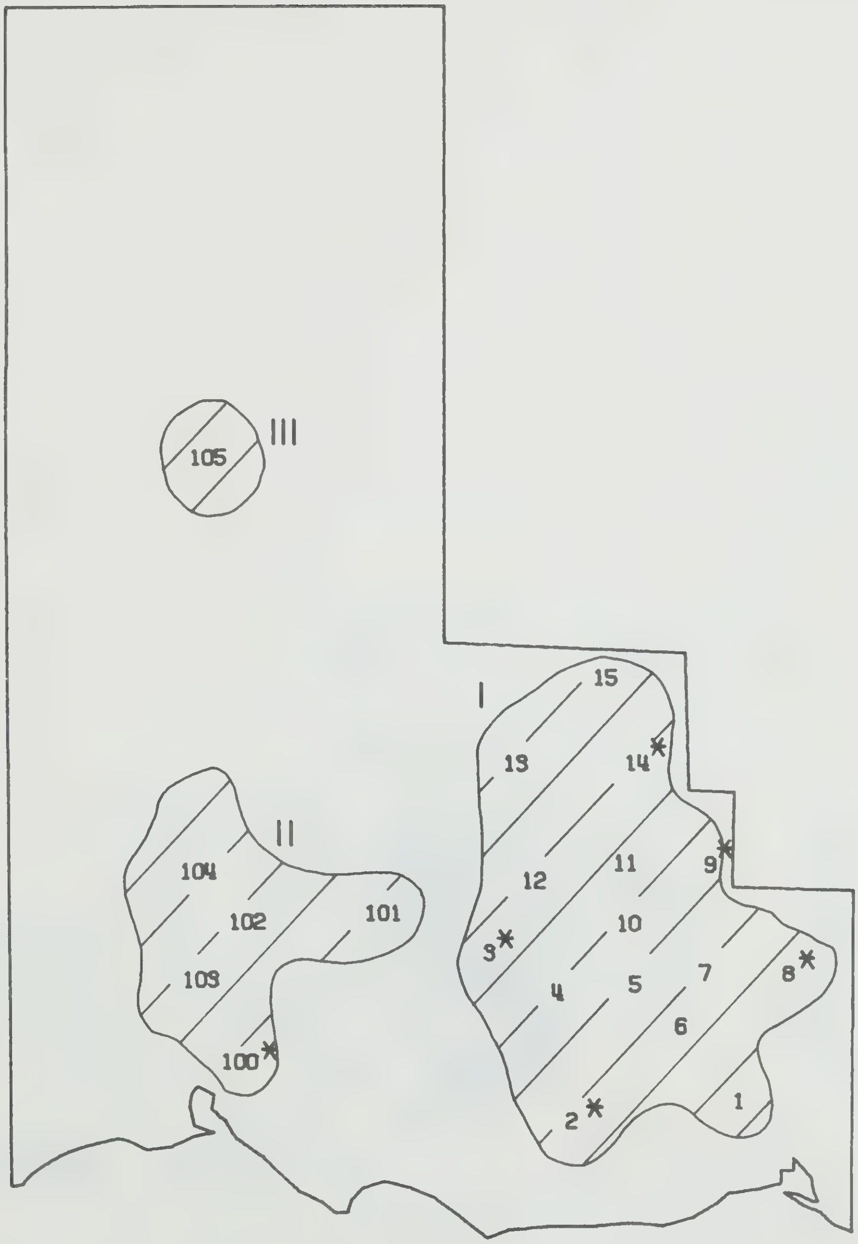


FIGURE 30. Census Tract Clusters for Oshawa

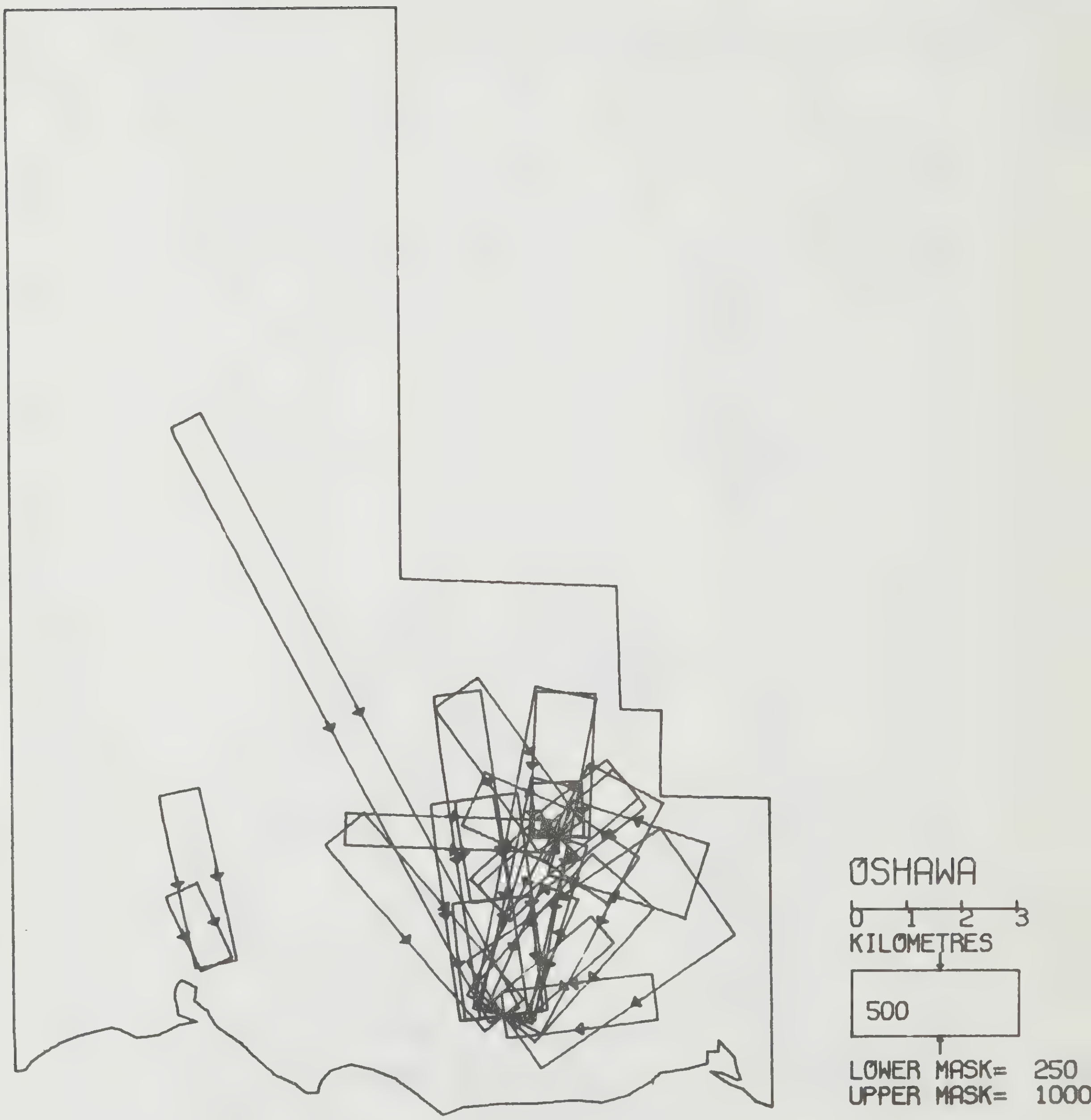


FIGURE 31. Major Home to Work Linkages for Oshawa

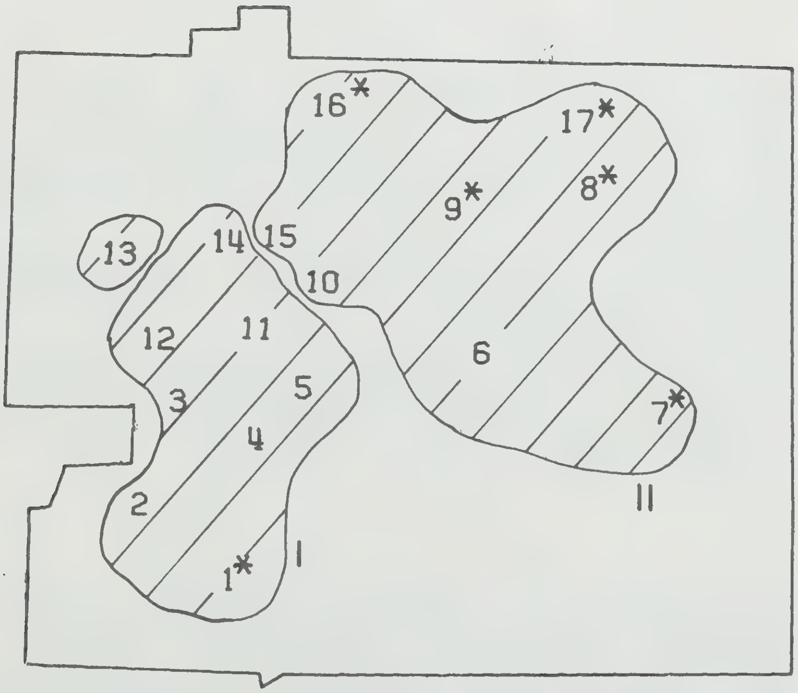


FIGURE 32. Census Tract Clusters for Sudbury

the CTs in the surrounding townships. The principal determinants of commuting in Sudbury are socio-economic differences along with the locations of mining employment. Figure 33 shows the major trip desires in Sudbury and demonstrates that significant linkage magnitudes exist only to the employment in the CBD.

Kitchener

Figure 34 shows four major census clusters for the Kitchener census area where the cluster boundaries follow closely the former municipal boundaries of Waterloo, Kitchener, Preston and Galt. The dendrogram in the appendix illustrates that the CTs in Preston and Galt cluster at about the same error sum of squares magnitude as the Kitchener CTs. A sharp increase in the error sum of squares occurs when the Kitchener and Waterloo census tracts are merged and a very high increase occurs when the Kitchener-Waterloo and Cambridge (Preston, Galt) clusters are merged. While manufacturing employment tended to dominate the economic base of the census area in 1971 there were major concentrations of non-manufacturing employment in the Kitchener CBD and at the University of Waterloo. As a result socio-economic factors have an important influence on commuting patterns in the census area as well as the municipal boundaries. Figure 35 illustrates the major trip desires in the Kitchener census area and shows fairly clearly the major commuter-shed sub-regions. Some of the linkages illustrated in Figure 35 are also due to the timing of development.

Windsor

Figure 36 identifies three clusters of census tracts in the Windsor census area. The CTs in cluster II to the east have strong linkages with CBD employment and to a lesser extent with the automobile manufacturing plant (CT 25). The CTs within cluster I are more strongly linked to employment in CT 25. The dendrogram shows that cluster III consists of 13 CTs only three of which are illustrated in Figure 36 with the remaining 10 being located in the townships outside of the city boundaries. The dendrogram illustrates that there is a significant increase in the error sum of squares when cluster I and II are merged and that there is a sharp increase when these are merged with cluster III. The principal determinants of trip linkages in Windsor are socio-economic factors, the timing of development and to a lesser extent the magnitude of employment in CT 25. Figure 37 illustrates the major trip desires in Windsor.

London

Figure 38 illustrates four clusters of census tracts for London. Cluster I embraces the CTs on the northeast fringe of the area which are newer residential areas with commuting linkages to the CBD (CT 22). Cluster II consists of the CTs to the east of the CBD which also have strong linkages with the CBD and to employment opportunities in CT 27. Cluster III embraces CTs to the west of the CBD and south of the river. These CTs also have a strong orientation to jobs in the CBD. Cluster IV is to the north of the river and these CTs tend to have strong linkages with employment at the

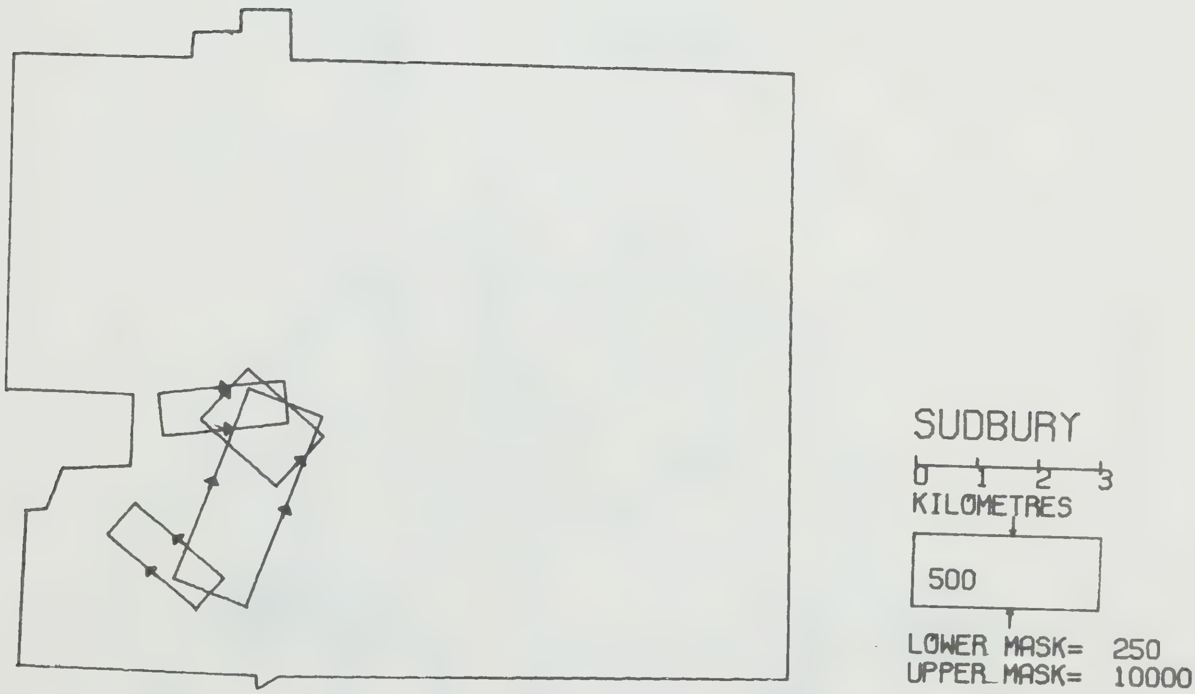


FIGURE 33. Major Home to Work Linkages for Sudbury

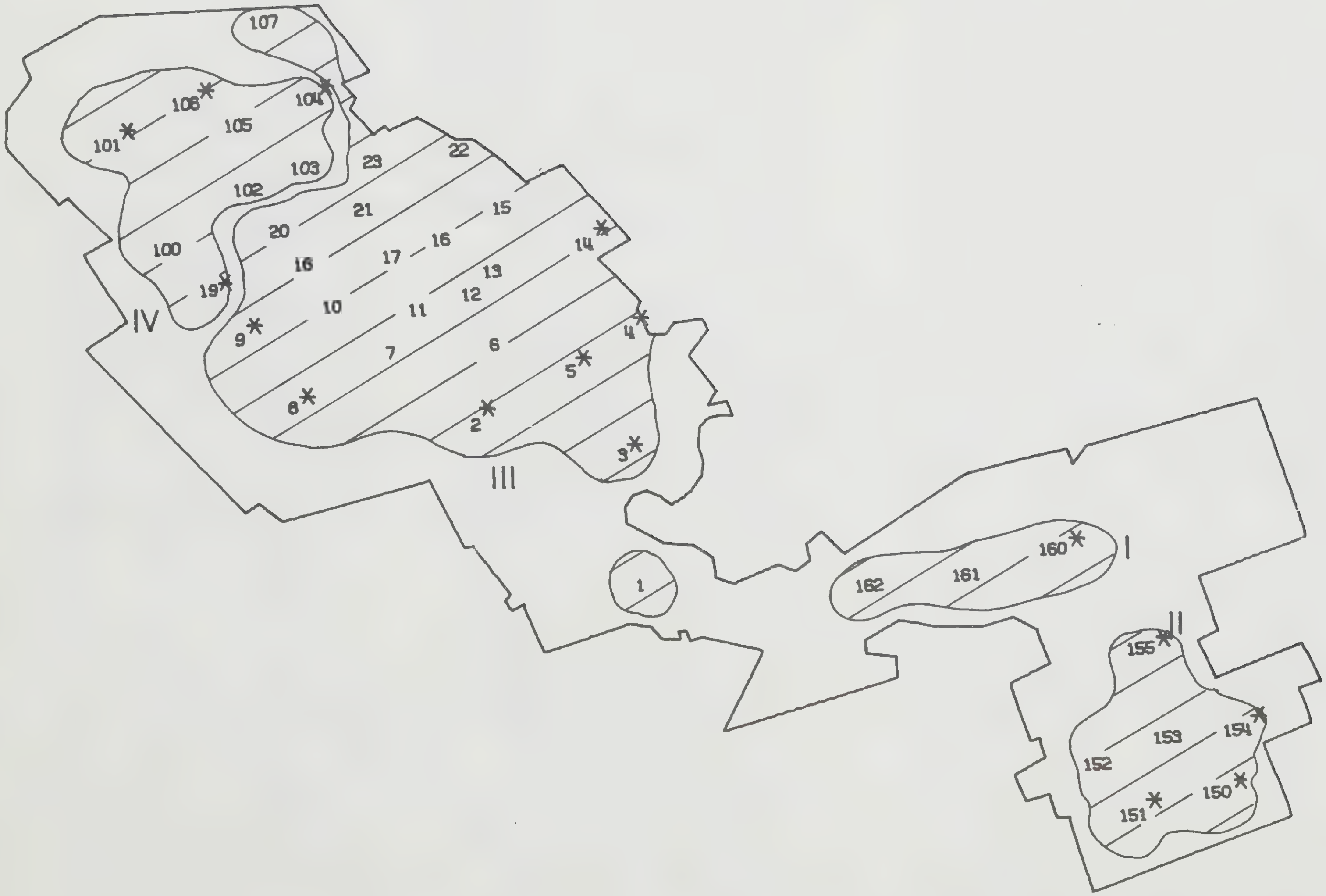


FIGURE 34. Census Tract Clusters for Kitchener Census Area

FIGURE 35. Major Home to Work Linkages for Kitchener Census Area

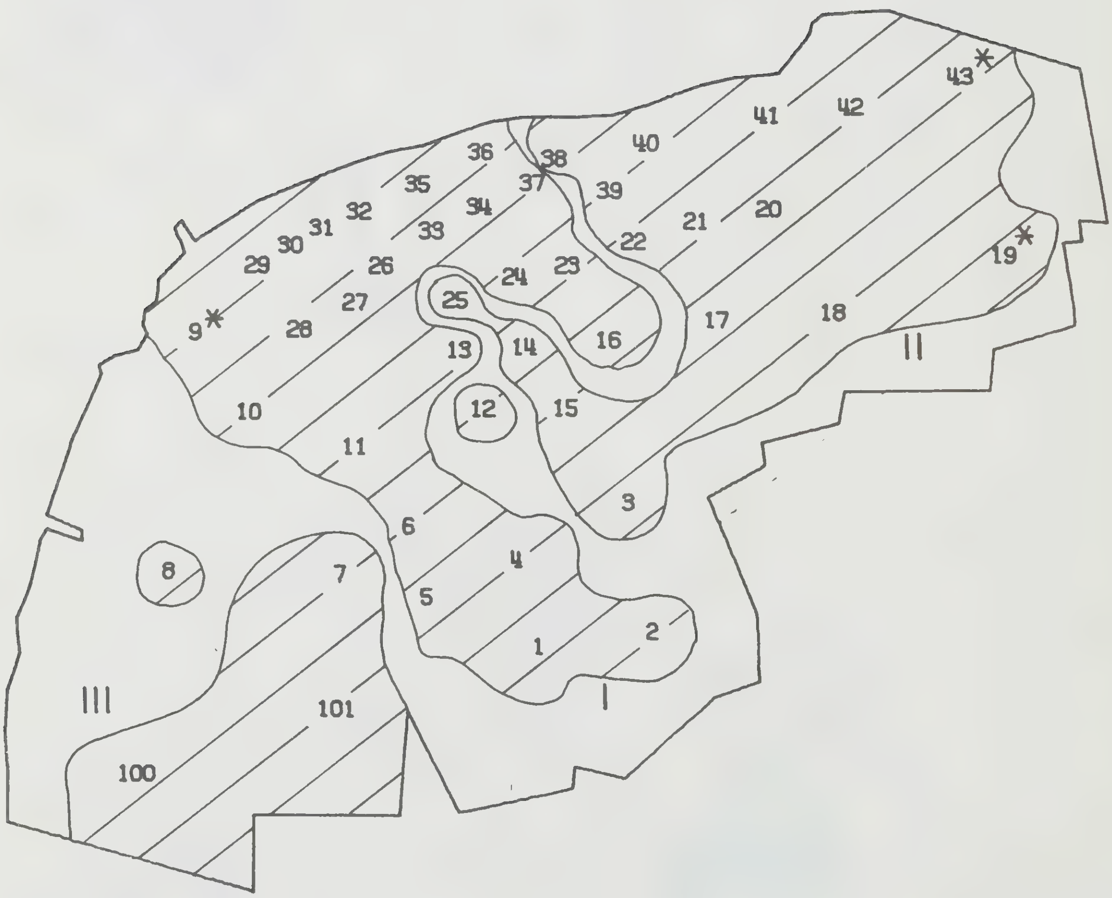


FIGURE 36. Census Tract Clusters for Windsor

FIGURE 37. Major Home to Work Linkages for Windsor

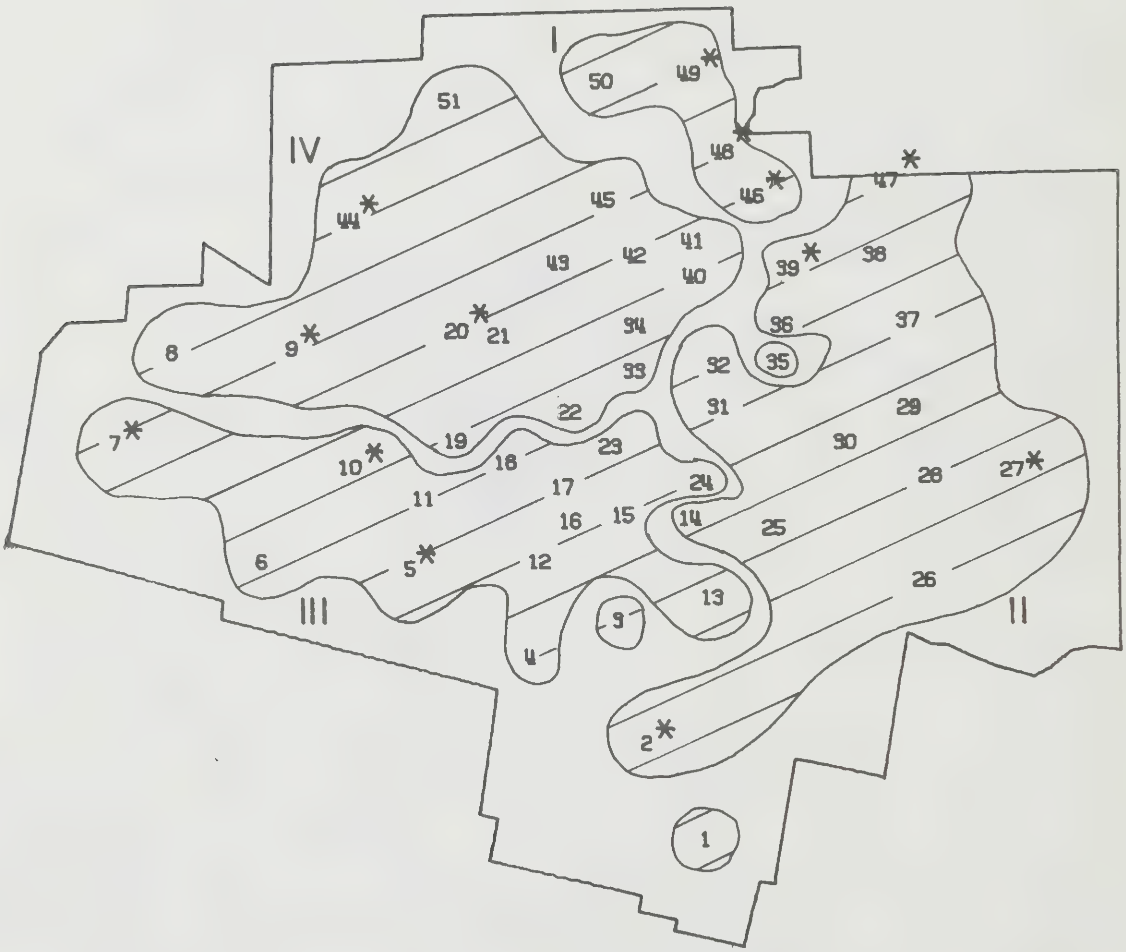


FIGURE 38. Census Tract Clusters for London

University of Western Ontario and to the CBD. The dendrogram shows that clusters I and II merge and clusters III and IV merge and that there is not a sharp increase in the error sum of squares when these merges occur. Figure 39 illustrates the major trip desires in London and demonstrates that commuting patterns are strongly dominated by the CBD employment and to a much lesser extent by the employment at the University of Western Ontario. The rivers and socio-economic factors also have an important impact on commuting patterns.

St. Catharines

Figure 40 illustrates the three major clusters of census tracts for the St. Catharines census area and this diagram shows that the cluster boundaries follow the municipal boundaries of St. Catharines, Niagara Falls and Welland. The dendrogram in the appendix illustrates that the CTs within each municipality cluster at low error sums of squares but that a very large increase in the error sum occurs when the three clusters are merged. Commuting patterns within each of the municipalities are determined principally by socio-economic factors. Figure 41 shows the trip desires for the area and these reflect the self-containment already mentioned in the dendrogram.

Hamilton

Figure 42 illustrates the census tract composition of five clusters for the Hamilton census area where three are located within Hamilton and two are within Burlington. The CTs in cluster I are located on the Niagara Escarpment and have strong linkages with the heavy industrial areas. Cluster II consists of CTs in the west end of Hamilton which have higher socio-economic status and these CTs have strong linkages with CBD employment and employment at McMaster Universtiy. Cluster II in the east end consists of lower status CTs which have strong linkages with the steel mill areas. The dendrogram in the appendix illustrates that a significant increase in the error sum of squares occurs when clusters I and II are merged with cluster III and that very large increases occur when the Hamilton clusters are merged with the Burlington clusters. In Burlington the CTs in cluster IV have strong linkages with Hamilton while those in cluster V are more strongly linked with employment opportunities in Burlington. Figure 43 illustrates the major trip desires in the Hamilton census area. Job concentrations, topographic features and socio-economic factors are the primary determinants of commuting patterns in the Hamilton census area.

Ottawa

Figure 44 illustrates eight census tract clusters for the Ottawa census area six of which are within the Ottawa-Carleton region and two of which are in Hull. Clusters II and IV embrace the older residential areas of Ottawa while the remaining six on the southside of the river consist primarily of the newer residential areas. The dendrogram in the appendix shows that clusters I and II which consist of CTs in a southerly corridor cluster, clusters III and IV in the east end cluster and clusters V and VI in the

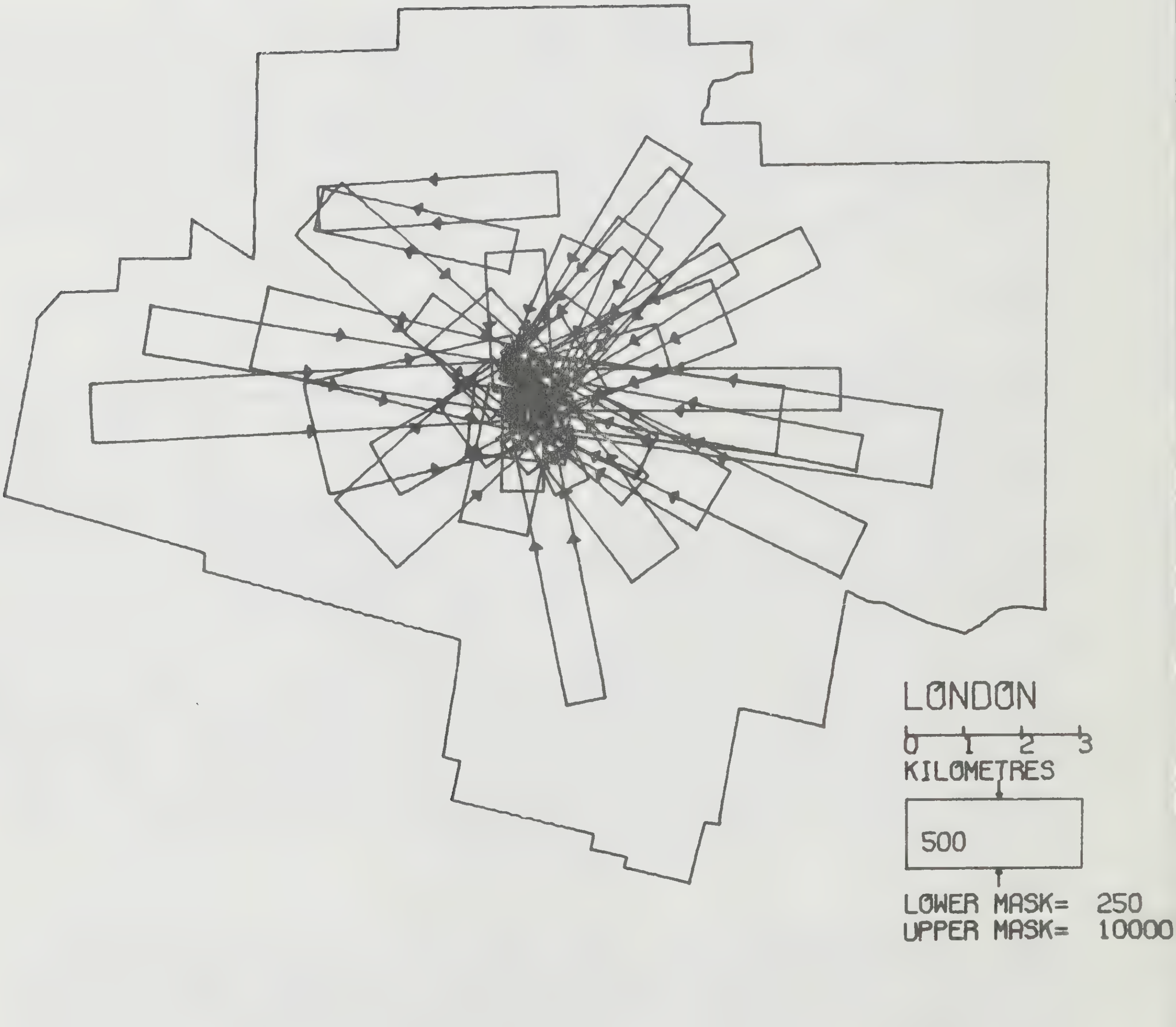


FIGURE 39. Major Home to Work Linkages for London

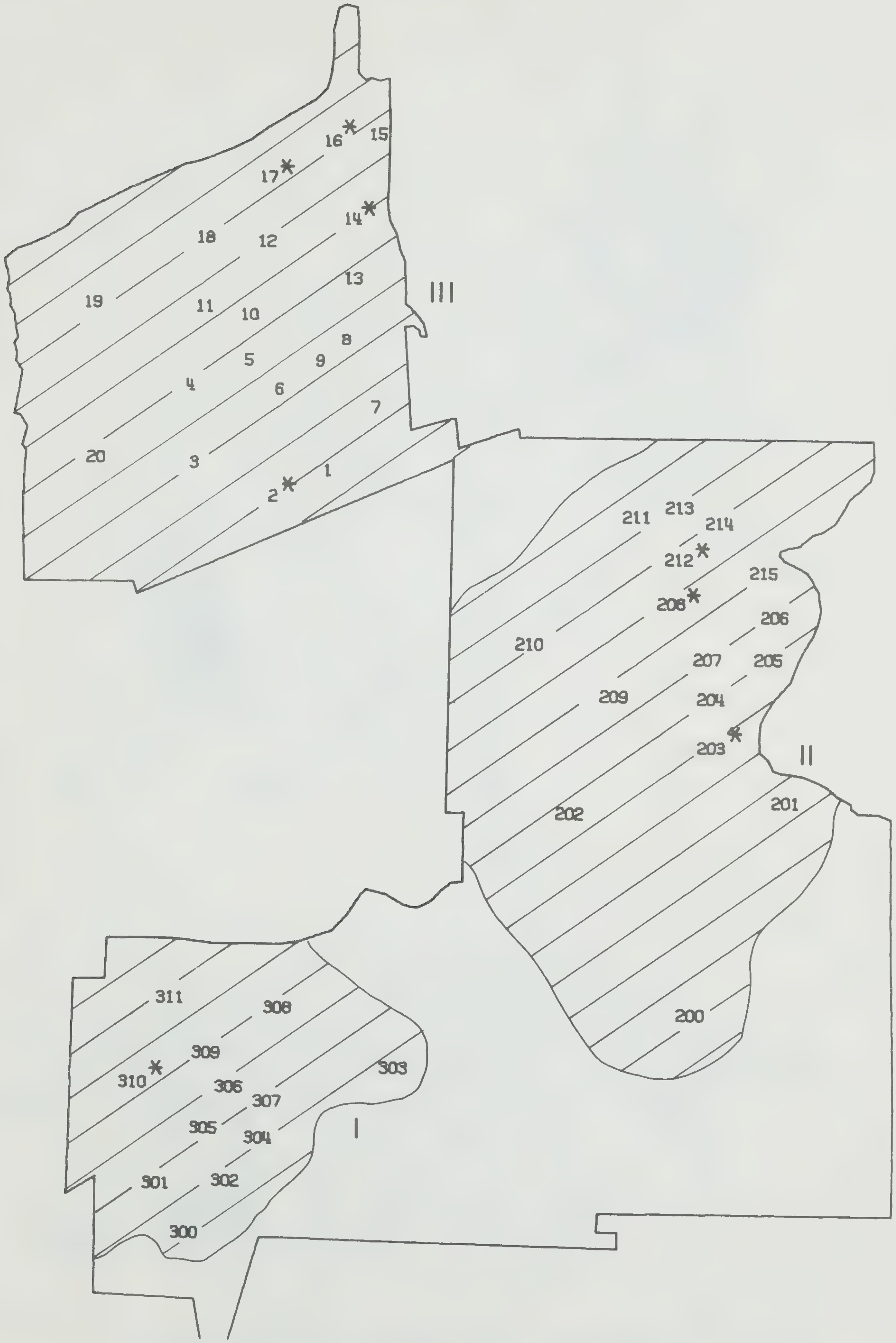


FIGURE 40. Census Tract Clusters for St. Catharines Census Area

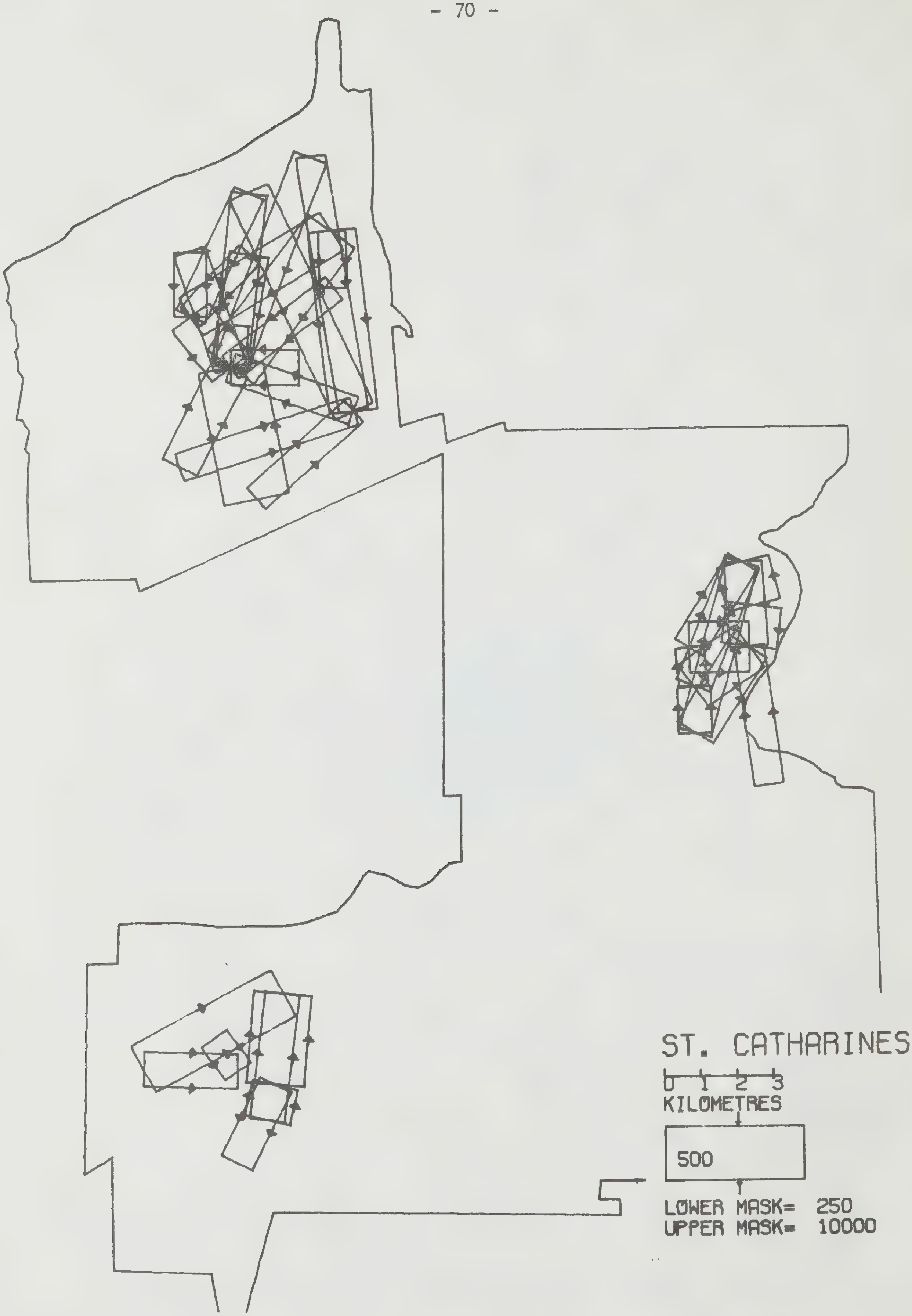


FIGURE 41. Major Home to Work Linkages for St. Catharines Census Area

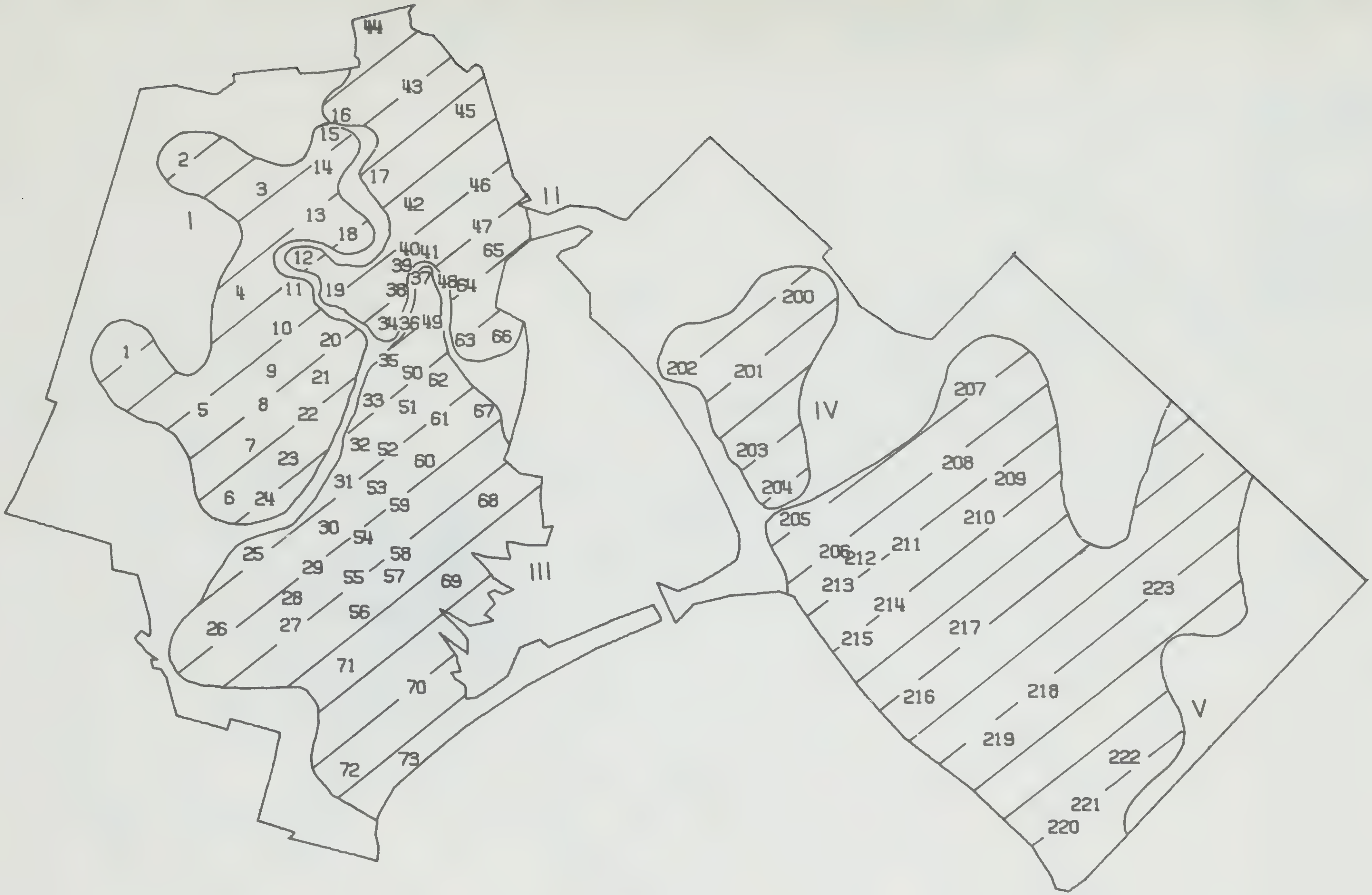


FIGURE 42. Census Tract Clusters for Hamilton Census Area

FIGURE 43. Major Home to Work Linkages for Hamilton Census Area

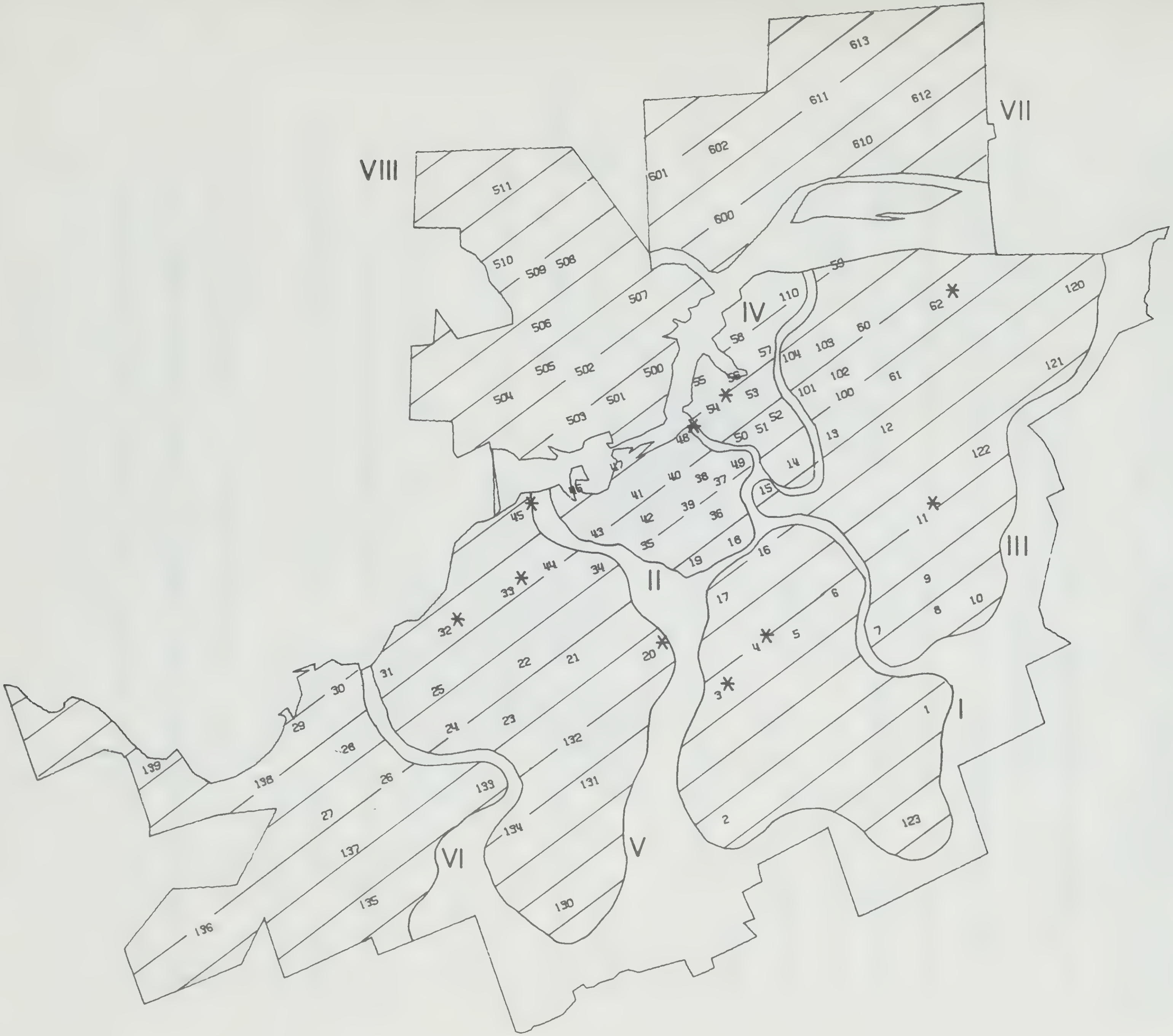


FIGURE 44. Census Tract Clusters for Ottawa Census Area

west end cluster. The next level of clustering illustrated by the dendrogram is the integration of clusters I, II, III and IV in the eastern half of the areas. There is a significant increase in the error sum of squares when all clusters on the southside of the river are joined and a sharp increase when clusters on both sides of the Ottawa River are joined. Commuting to the central area tends to dominate the commuting pattern as shown in Figure 45. In addition the timing of development has a strong influence as do socio-economic characteristics and topographic features.

2.4 Some Generalizations About Commuting Patterns

The previous paragraphs have outlined the census tract composition of the major clusters of residential areas that have similar destination patterns. The principal determinants of these commuting patterns have been highlighted for each census area under each of the five broad categories identified in Section 2.3. Table 4 summarizes the major determinants of commuting in each of the fifteen census areas in terms of each of these areas.

Seven of the census areas as defined by Statistics Canada embrace two or more separate municipalities. In almost all cases major cluster boundaries were coincident with the municipal boundaries indicating that commuting patterns tend to be self-contained within municipalities. Thunder Bay, Kitchener, St. Catharines and Hamilton provide very good examples of this type of effect.

Topographic features such as rivers also have an important influence on commuting patterns. The best examples are provided by London, Hamilton and Ottawa. Ottawa is a special case in that the Ottawa River is also a provincial boundary.

One of the most important influences on commuting patterns is the timing of development where residential areas and employment zones that expanded during the same time period tend to have strong linkages.

FIGURE 45. Major Home to Work Linkages for Ottawa Census Area

TABLE 4. Major Determinants of Residential Zone Commuting Clusters

Census Area	Multi-Community Composition of Census Area	Topographic Features	Timing of Development	Socio-Economic Factors	Specific Employment Concentrations
Guelph			✓	✓	
Peterborough			✓	✓	
Sarnia			✓	✓	✓
Brantford	✓	✓	✓	✓	
Sault Ste. Marie				✓	✓
Kingston		✓	✓	✓	
Thunder Bay	✓			✓	
Oshawa	✓		✓	✓	✓
Sudbury				✓	✓
Kitchener	✓		✓	✓	
Windsor			✓	✓	✓
London		✓	✓	✓	✓
St. Catharines	✓			✓	
Hamilton	✓	✓	✓	✓	✓
Ottawa	✓	✓	✓	✓	✓

The best example is provided by Guelph where the residential areas and employment zones that grew rapidly during the 1966-1971 period were all located on the periphery. The commuting patterns that emerged as a consequence of this phase of development have a distinctly different character to the centrally focussed patterns that developed earlier.

Perhaps the strongest influence on the commuting structure of a community are the socio-economic factors that influence dwelling unit location decisions and employment location possibilities. The importance of socio-economic factors is analyzed in more detail in subsequent chapters of this report.

The final factor identified in Table 4 is the impact that specific employment concentrations have on commuting patterns. Communities which have large concentrations of employment in one area either because of one industry or a collection of employers in that area develop particular types of commuting patterns that must be recognized in any modelling attempts. Sarnia, Sault Ste. Marie, Oshawa, Sudbury, Windsor and Hamilton provide examples of areas with strong concentrations of industries in particular areas and the commuting patterns to these areas tend to dominate the home to work linkages patterns. London and Ottawa are examples in which employment in the tertiary sector dominates the economic base of the community and this employment type tends to be concentrated in the CBD.

The five determinants identified above and the associated clusters described in the previous sections provide a basis for identifying calibration sub-regions within each of the census areas. The particular sub-regions used in the calibration of multi-parameter gravity models are described later in this report.

Some understanding of the heterogeneity in commuting patterns in each of the census areas is provided by Table 5 in which the maximum value of error sum of squares is reported for each census area along with the number of spatial linkages in each area. Clearly this magnitude will increase with increasing size of urban area but deviations from the general trend may be clearly identified. For example in the smaller urban areas the higher than average maximum error sum of squares for Sarnia, Sault Ste. Marie and Kingston may be noted. In the next size group Thunder Bay, Oshawa and Kitchener provide examples of census areas in which commuting heterogeneity leads to sharp increases in the error sums of squares in the final stages of the clustering process. Finally in the three largest census areas analyzed the error sums of squares also increased sharply in the final clustering stages primarily because residential zones from separate municipalities were being forced together.

TABLE 5. Maximum Value of Error Sum of Squares for Ontario Census Areas

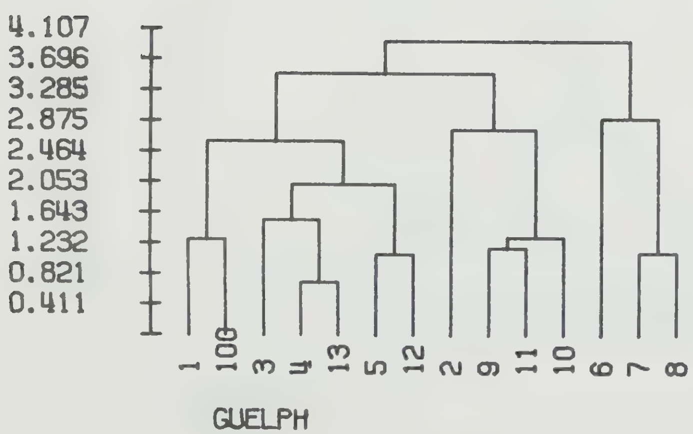
Census Area	Number of Home-Work Linkages	Error Sum of Squares
Guelph	19,990	3.9
Peterborough	20,400	5.3
Sarnia	23,760	6.3
Brantford	24,195	5.6
Sault Ste. Marie	24,740	6.5
Kingston	26,770	6.8
Thunder Bay	33,130	15.6
Oshawa	33,740	10.9
Sudbury	45.580	8.5
Kitchener	81,095	22.0
Windsor	74,455	10.3
London	97,195	8.1
St. Catharines	91,815	21.4
Hamilton	152,265	20.0
Ottawa	201,015	19.2

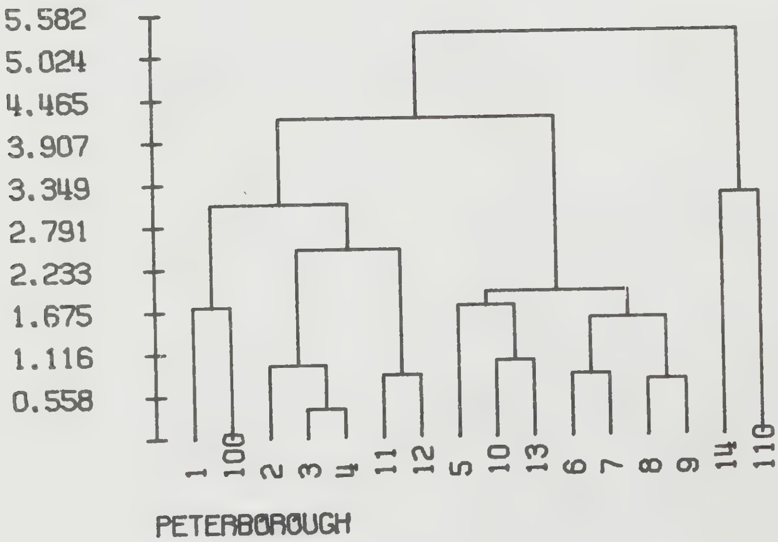
APPENDIX TO CHAPTER 2

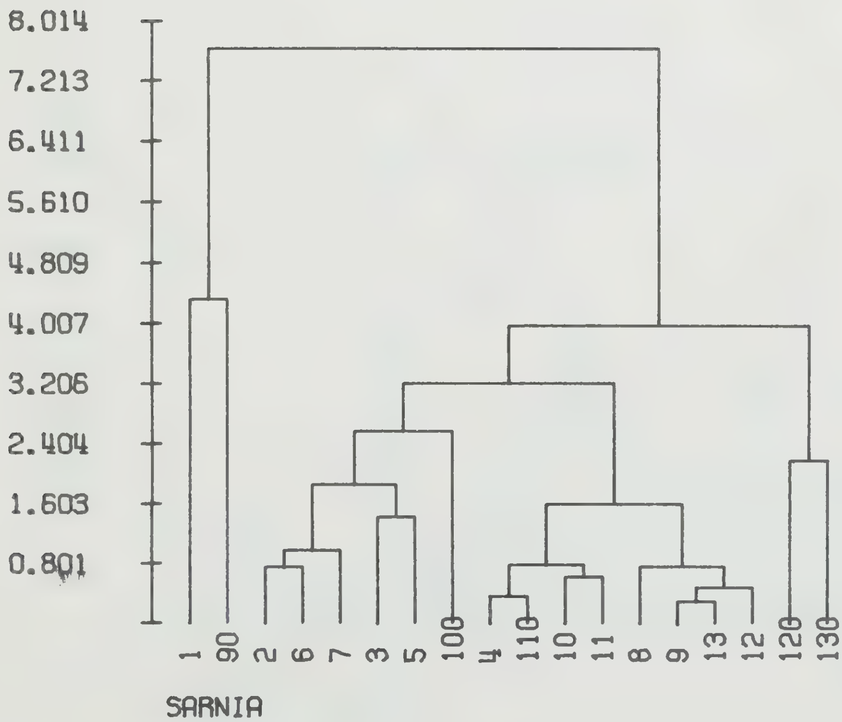
DENDROGRAMS FOR ONTARIO CENSUS AREAS

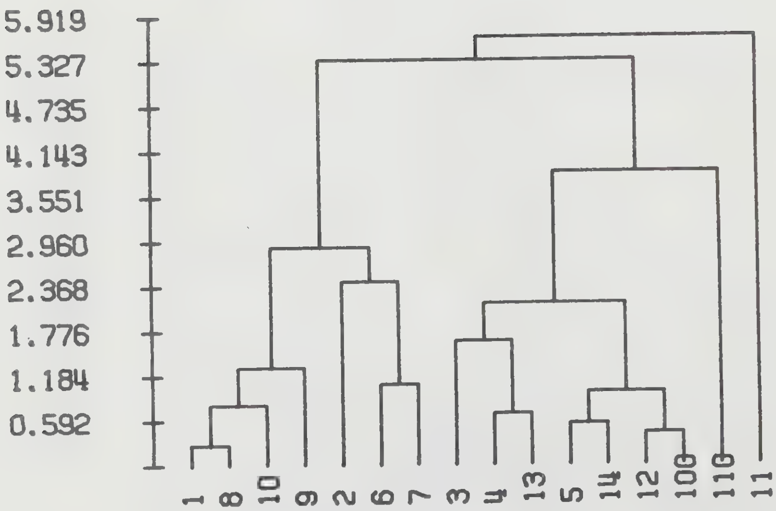
The dendrograms for the fifteen Ontario census areas are included in this appendix. These dendrograms show the the hierarchical structure of residential census tracts which have been grouped on the basis of similarity in destination zone characteristics.

The dendrograms are all plotted to the same error sum of squares scale along the ordinate in order to illustrate the degree of heterogeneity in commuting patterns. Error sums of squares are not directly comparable since the magnitude does increase with the number of census tracts used. The official census tracts only are included in the maps illustrated in this appendix. All census tracts are included in the dendrograms.

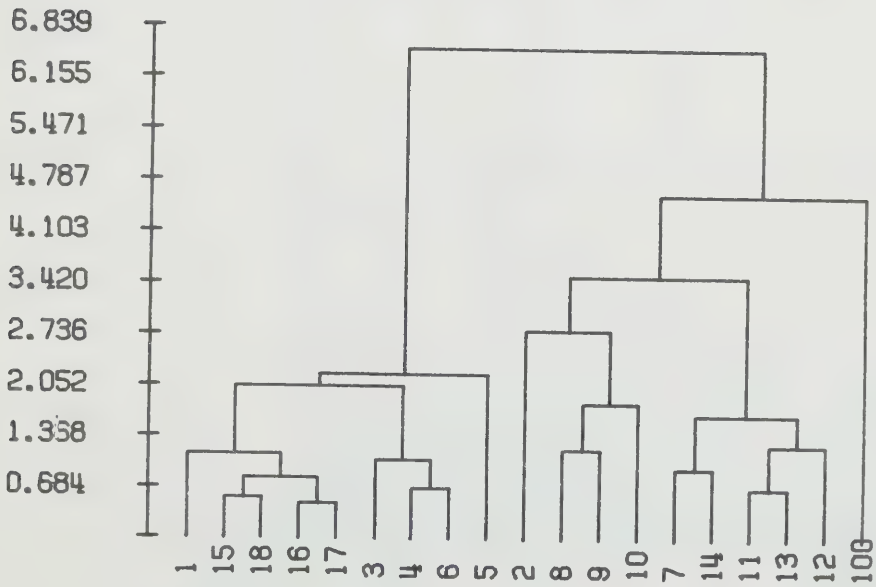

4.107
3.696
3.285
2.875
2.464
2.053
1.643
1.232
0.821
0.411
1 100 3 4 13 5 12 2 9 11 10 6 7 8
GUELPH


5.582
5.024
4.465
3.907
3.349
2.791
2.233
1.675
1.116
0.558
1
100
2
3
4
11
12
5
10
13
6
7
8
9
14
110
PETERBOROUGH

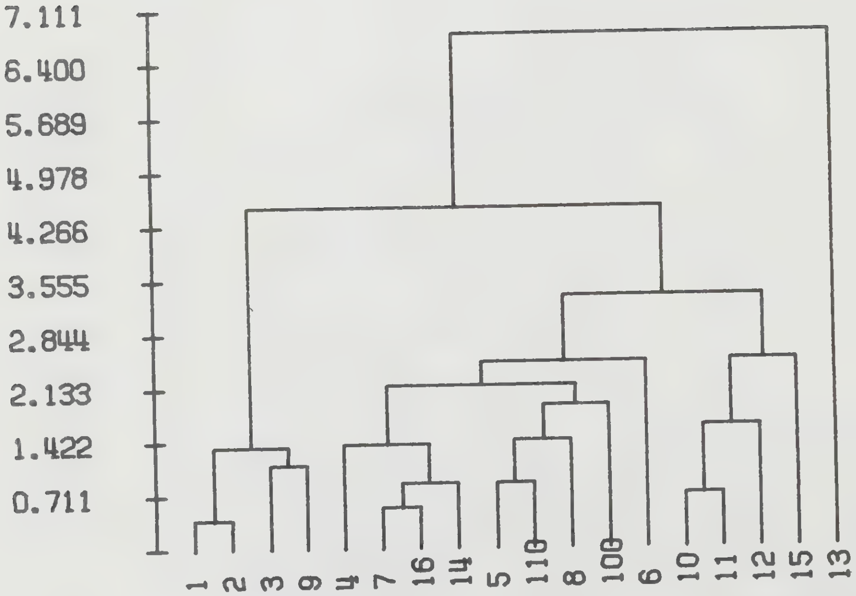

8.014
7.213
6.411
5.610
4.809
4.007
3.206
2.404
1.603
0.801
1
90
2
6
7
3
5
100
4
110
10
11
8
9
13
12
120
130
SARNIA


5.919
5.327
4.735
4.143
3.551
2.960
2.368
1.776
1.184
0.592
1
8
10
9
2
6
7
3
4
13
5
14
12
100
110
11

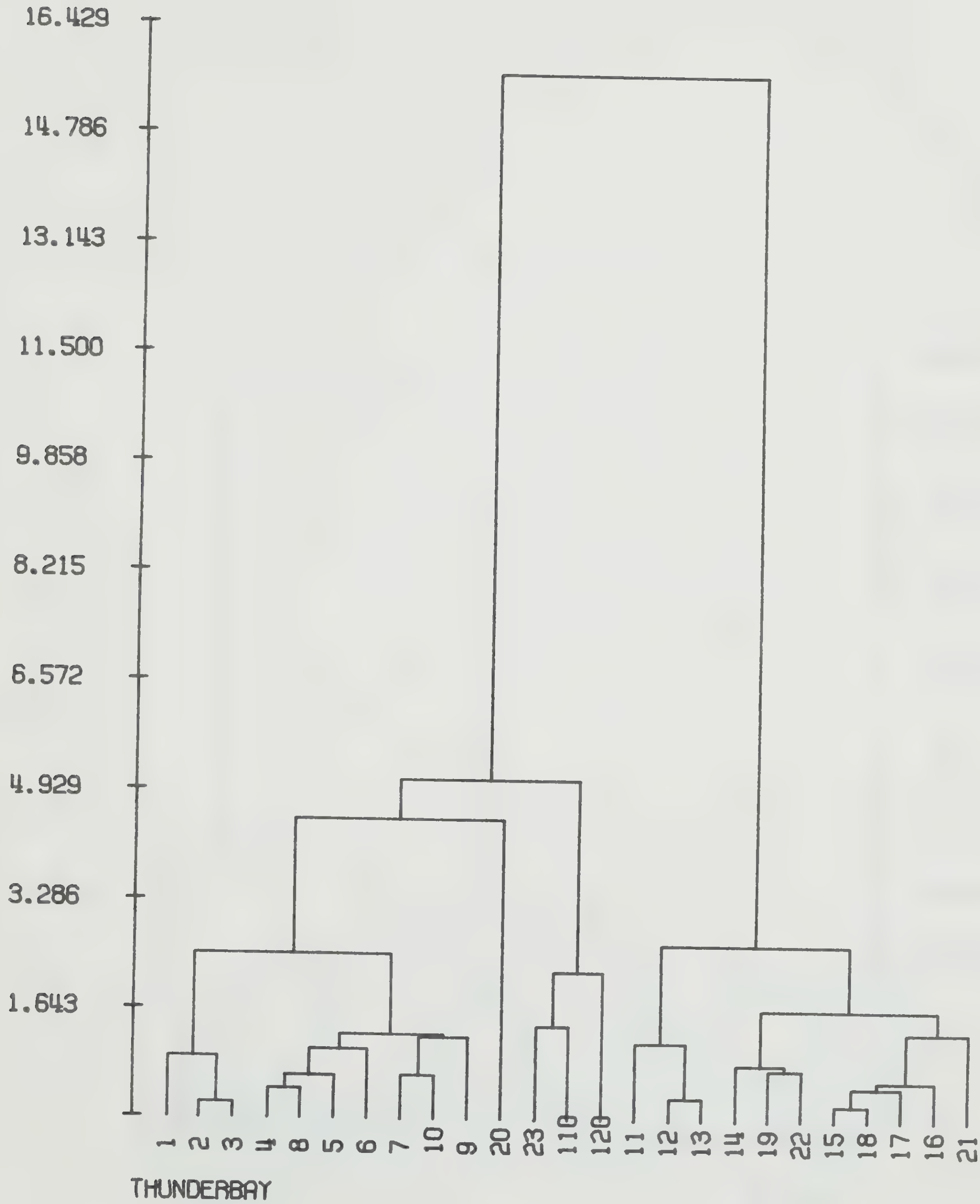
BRANTFØRD

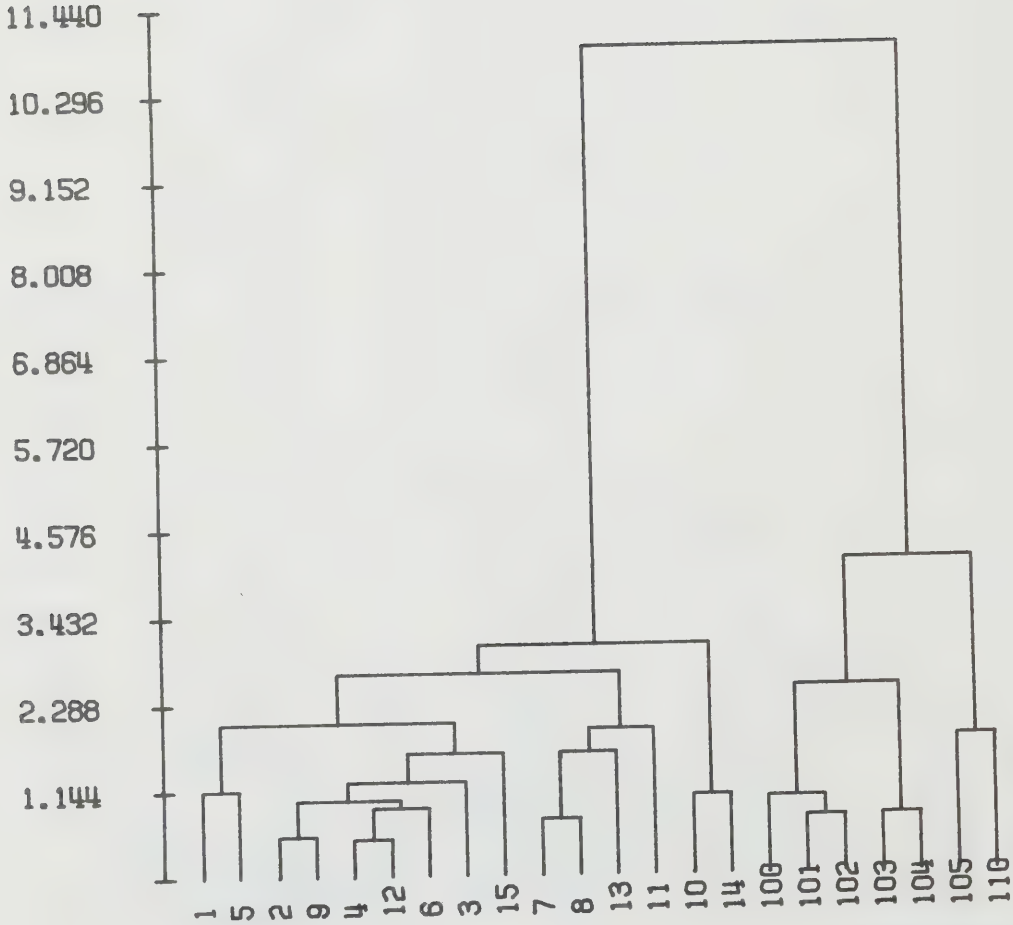


SAULT STE. MARIE


7.111
6.400
5.689
4.978
4.266
3.555
2.844
2.133
1.422
0.711
1
2
3
9
4
7
16
14
5
110
8
100
6
10
11
12
15
13

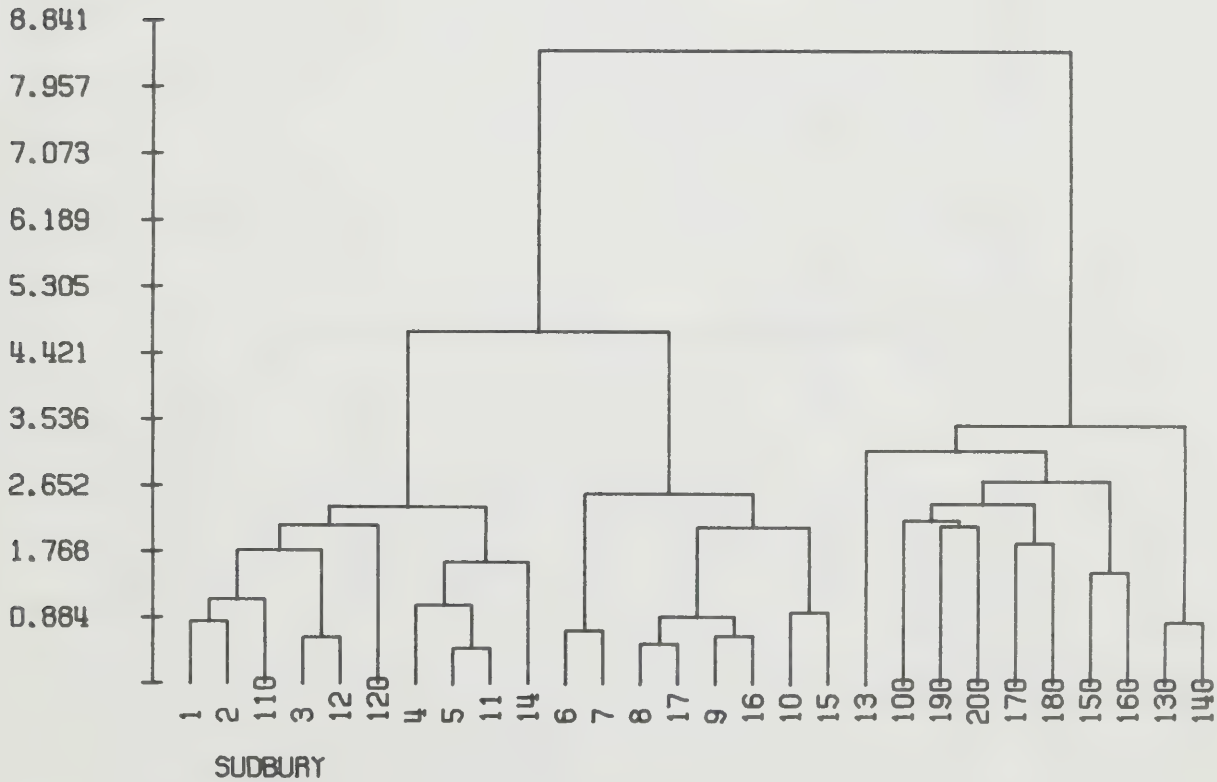
KINGSTØN

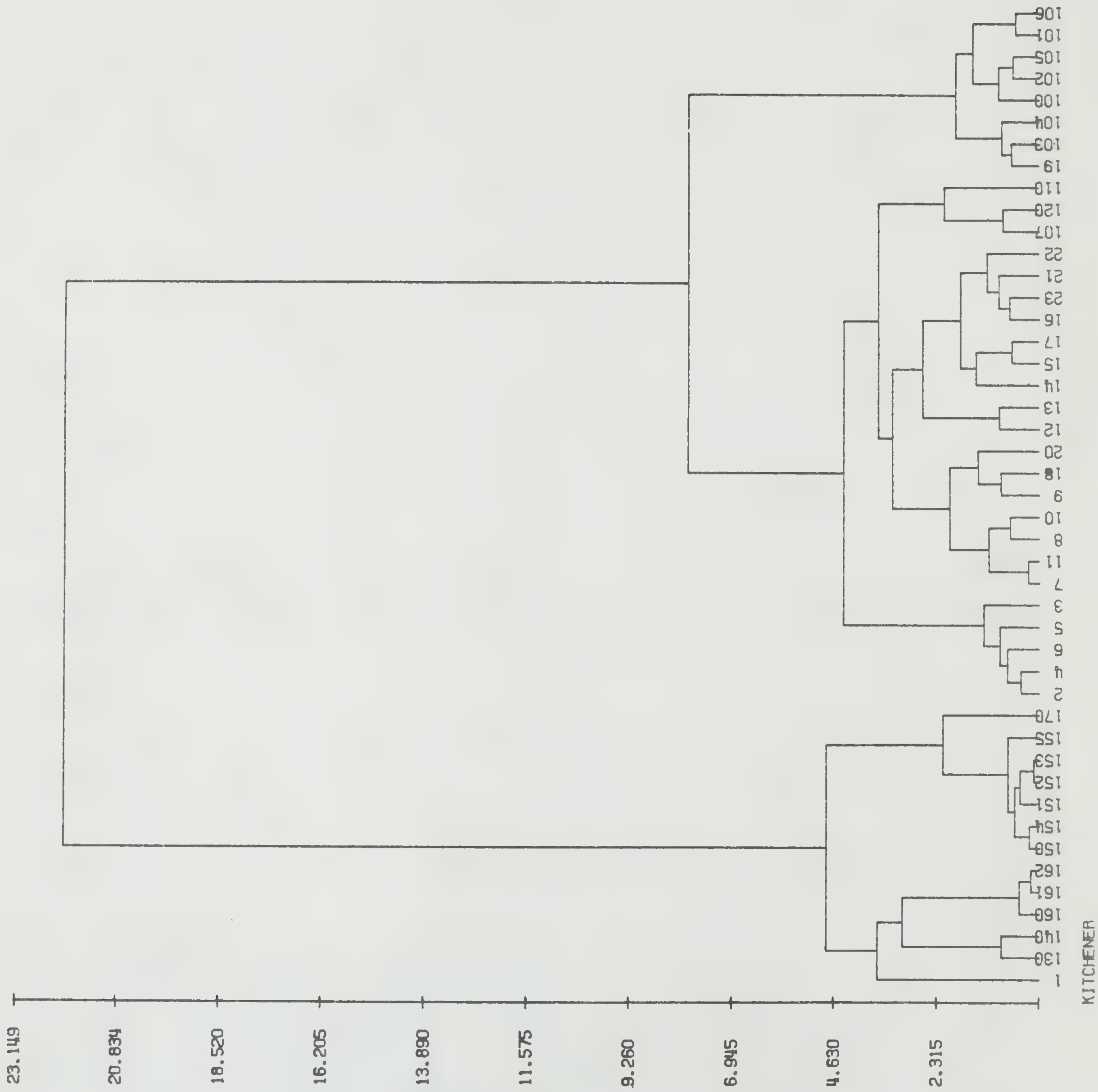

16.429
14.786
13.143
11.500
9.858
8.215
6.572
4.929
3.286
1.643
1 2 3 4 8 5 6 7 10 9 20 23 110 120 11 12 13 14 19 22 15 18 17 16 21
THUNDERBAY

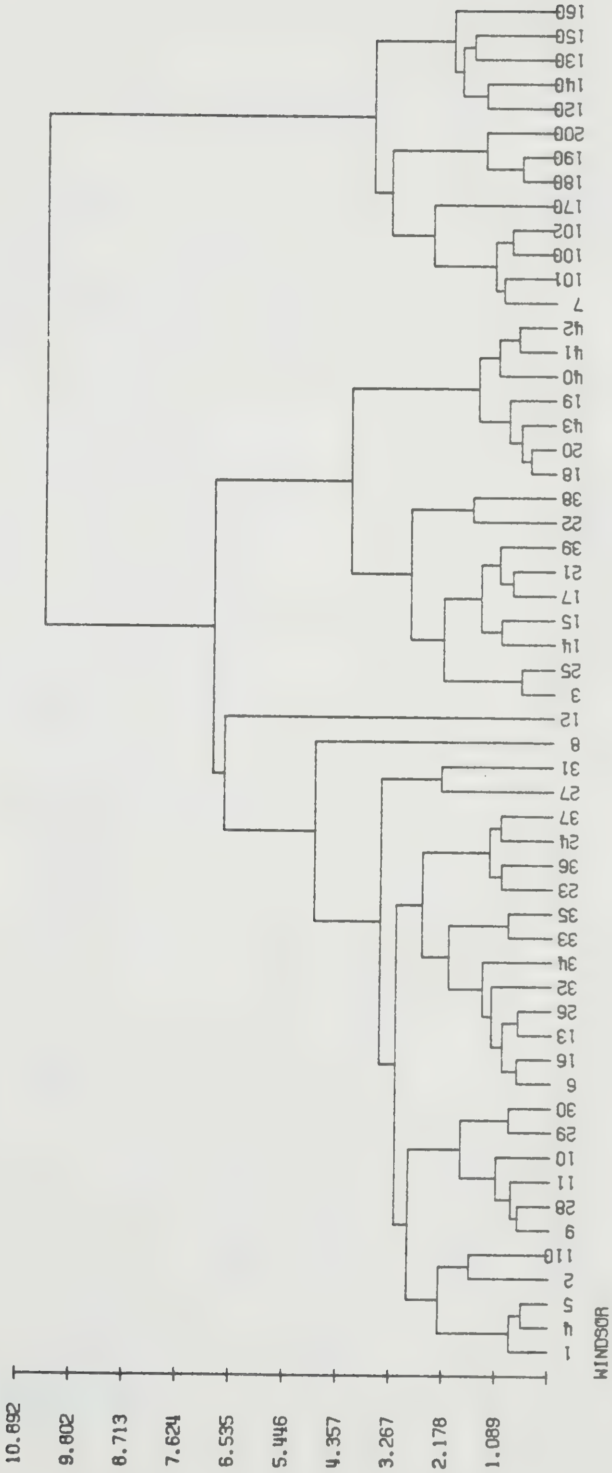


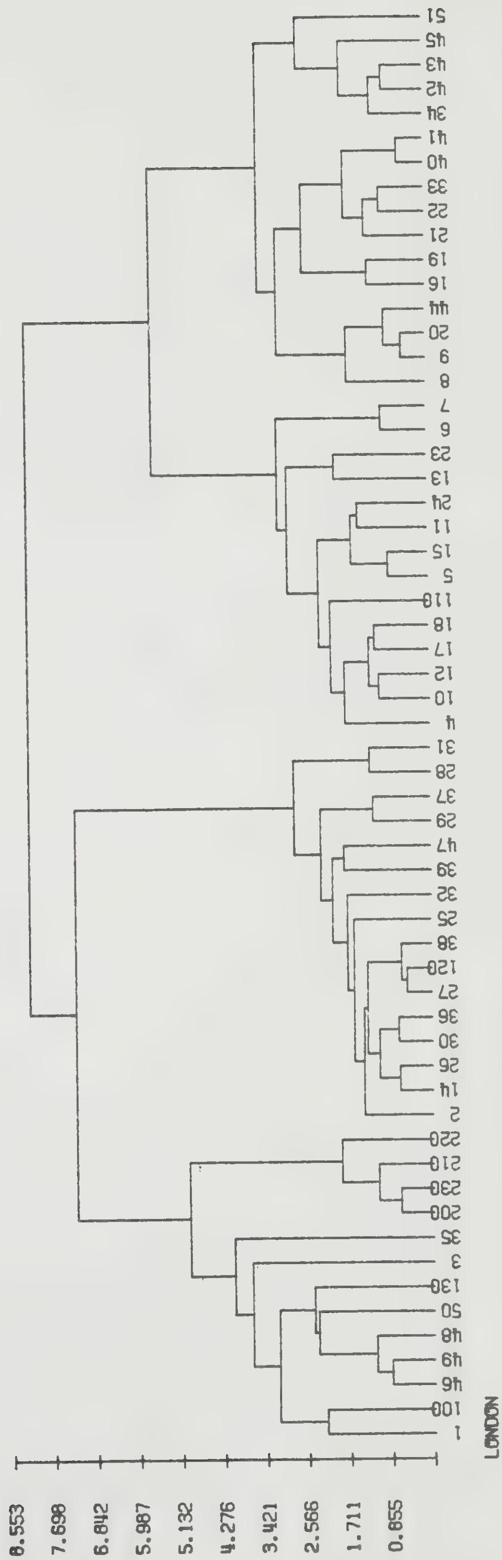
11.440
10.296
9.152
8.008
6.864
5.720
4.576
3.432
2.288
1.144
1 5 2 9 4 12 6 3 15 7 8 13 11 10 14 100 101 102 103 104 105 110

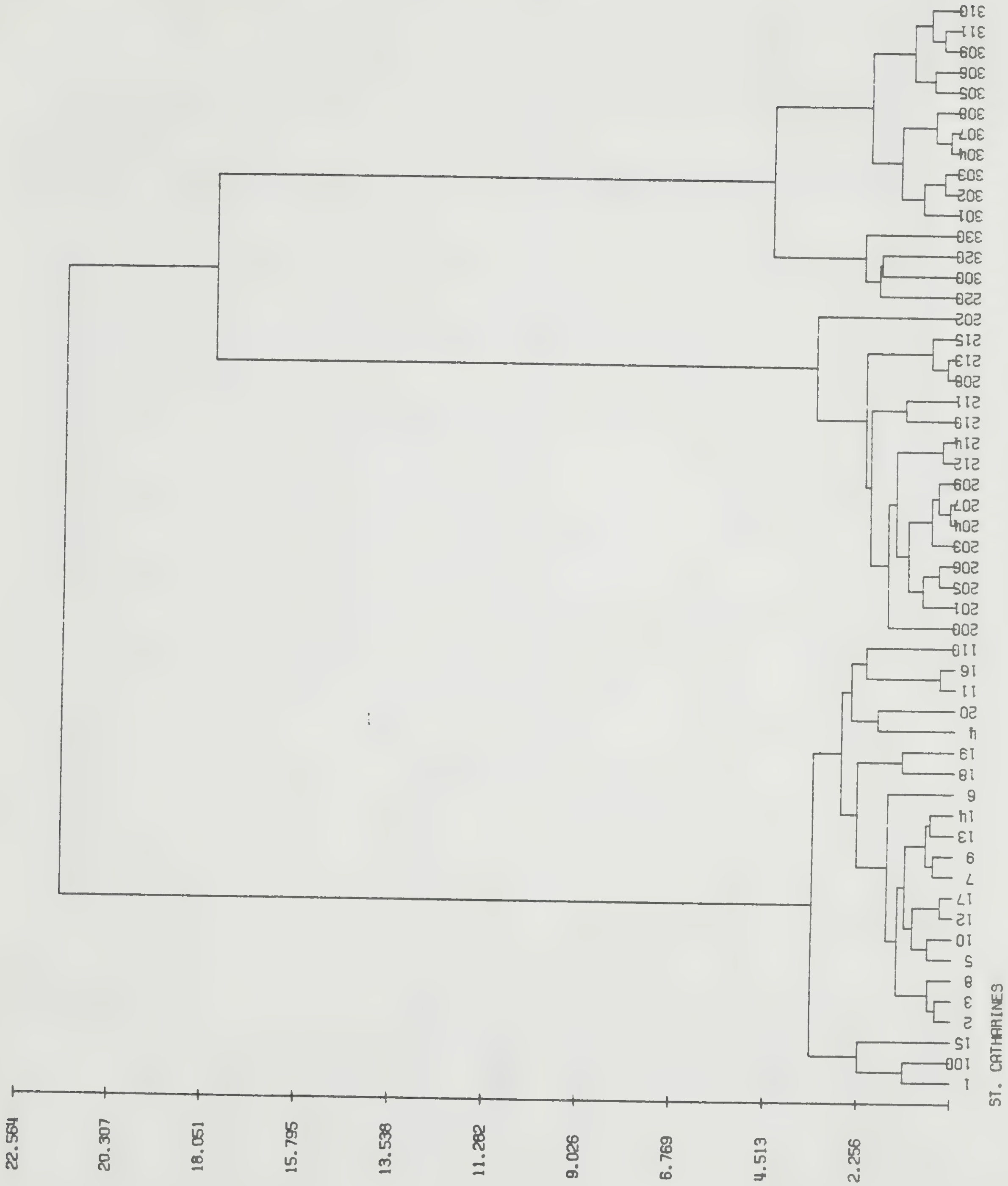
OSHAWA

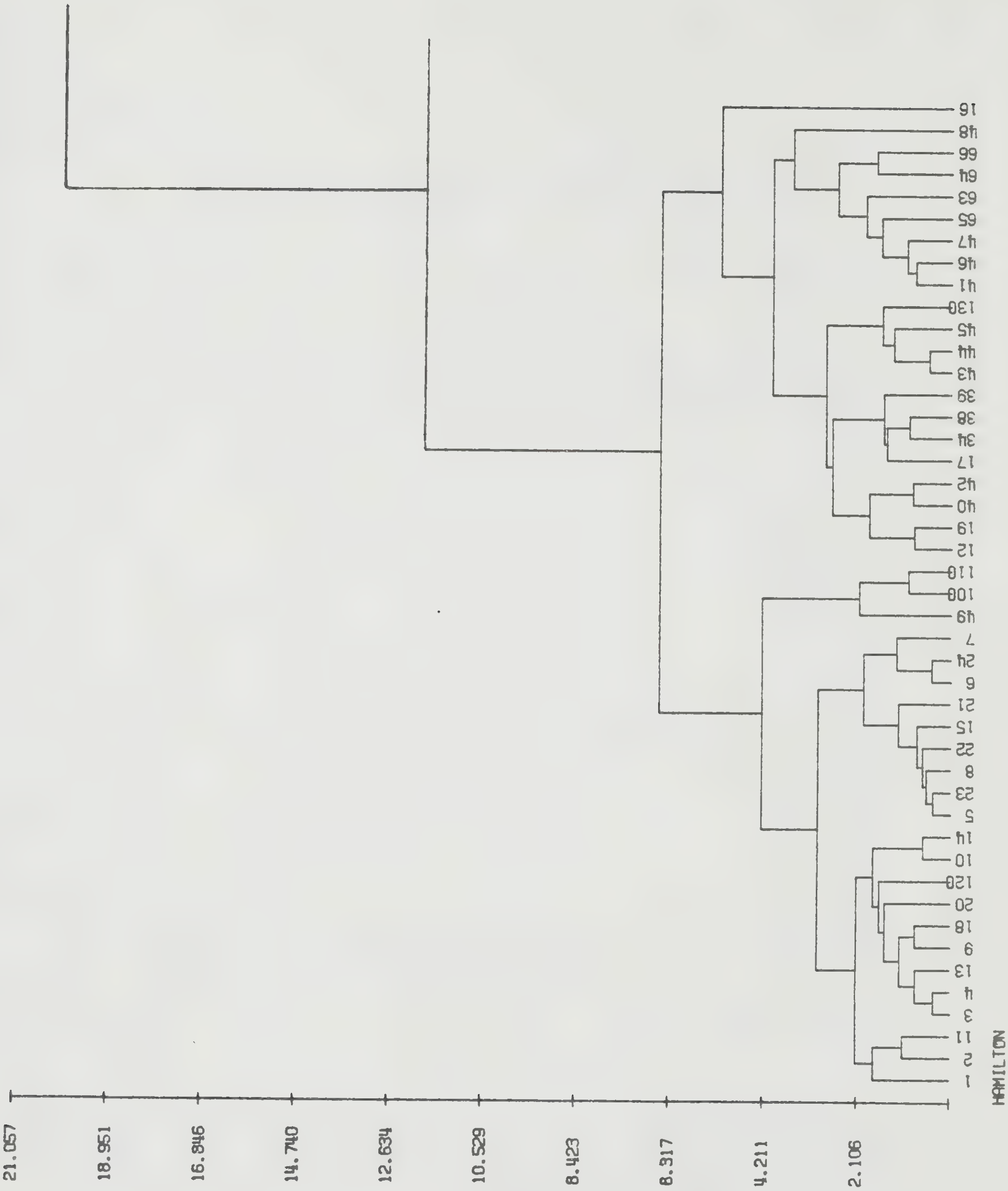

8.841
7.957
7.073
6.189
5.305
4.421
3.536
2.652
1.768
0.884
1 2 110 3 12 120 4 5 11 14 6 7 8 17 9 16 10 15 13 100 190 200 170 180 150 160 130 140
SUDBURY

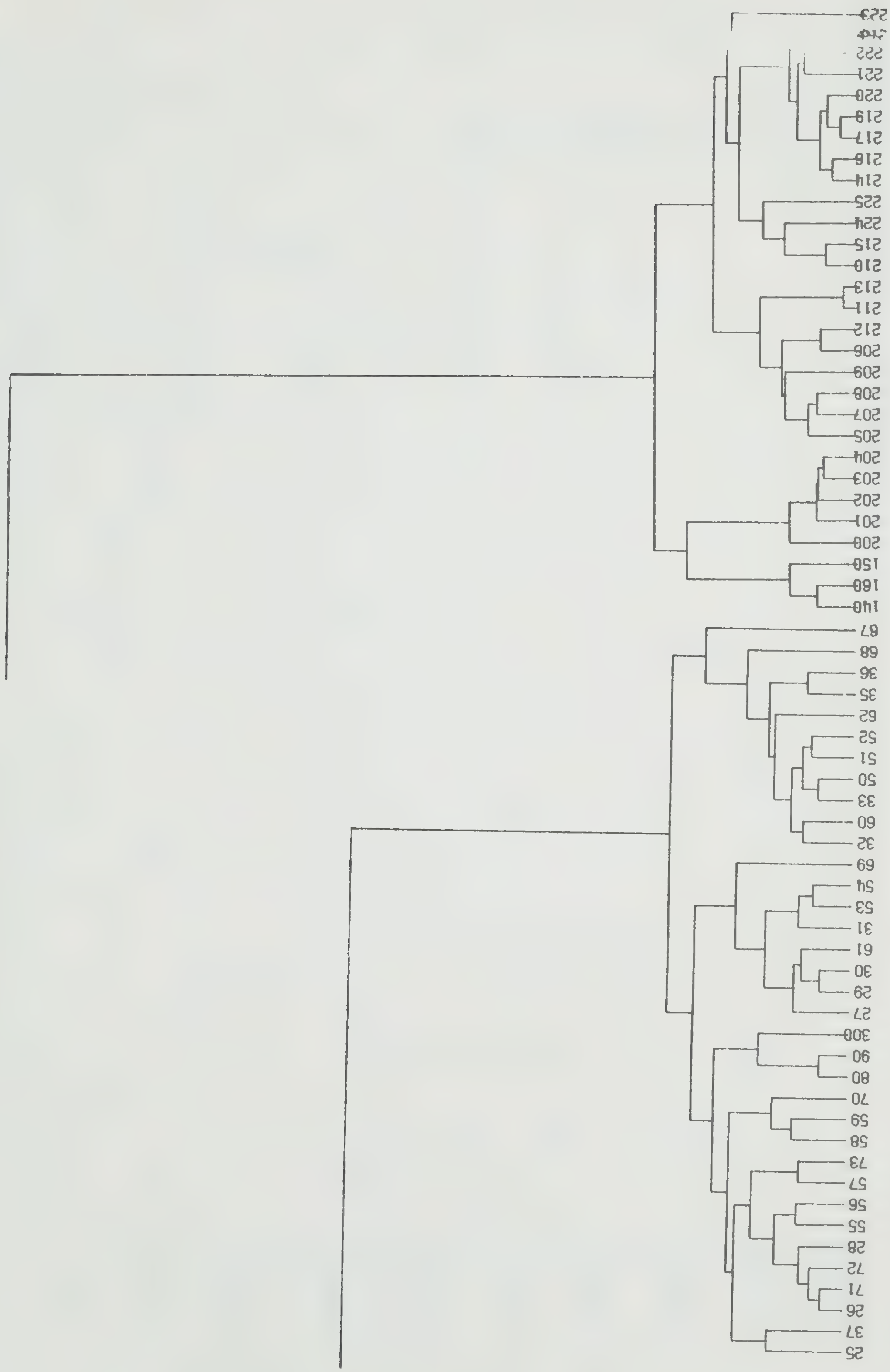

106
101
105
102
100
104
103
19
110
120
107
22
21
23
16
17
15
14
13
12
20
18
9
10
8
11
7
3
5
6
4
2
170
155
153
152
151
154
150
162
161
160
140
130
1
KITCHENER
23.149
20.834
18.520
16.205
13.890
11.575
9.260
6.945
4.630
2.315

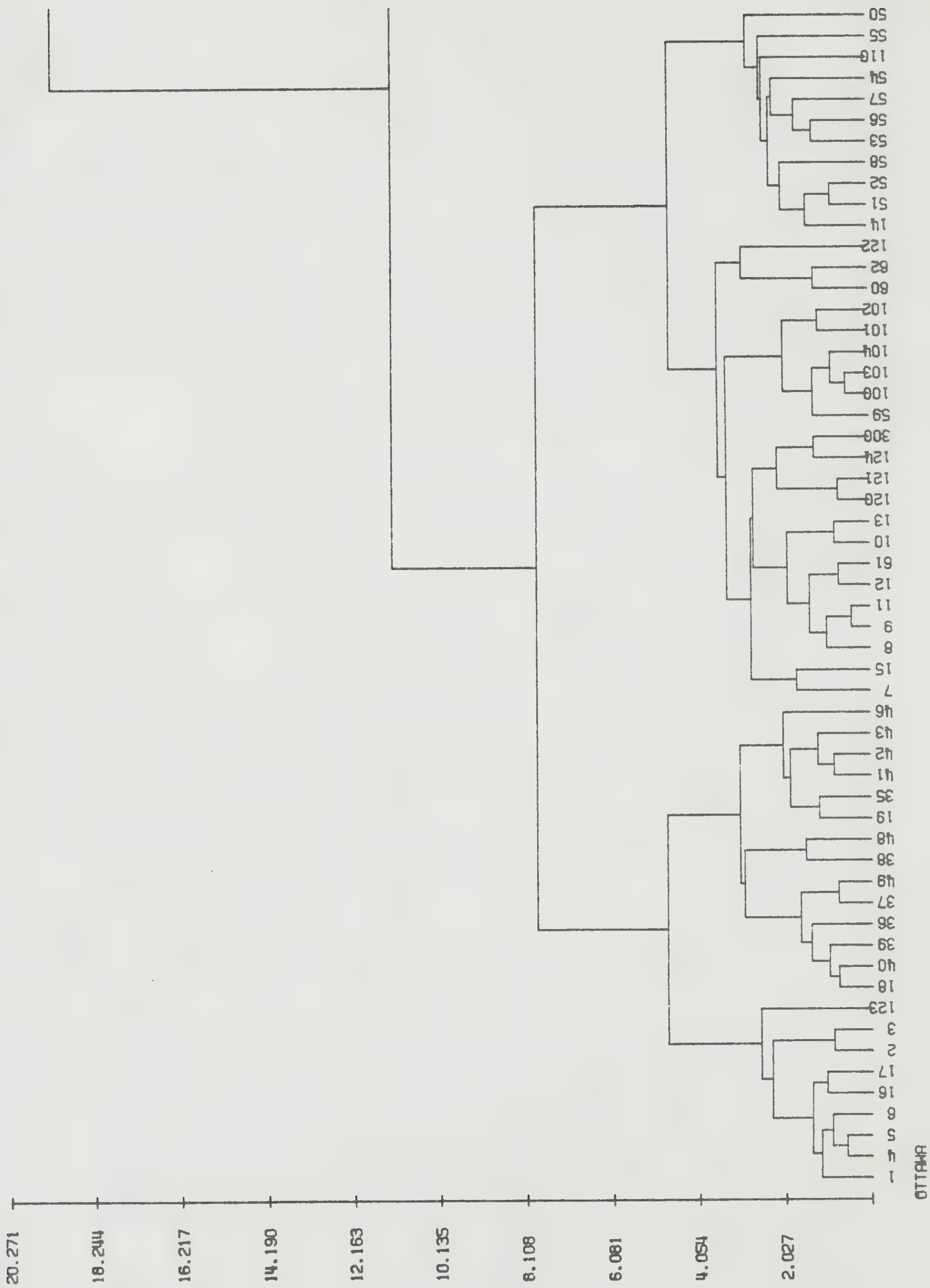

10.892
9.802
8.713
7.624
6.535
5.446
4.357
3.267
2.178
1.089
WINDSOR

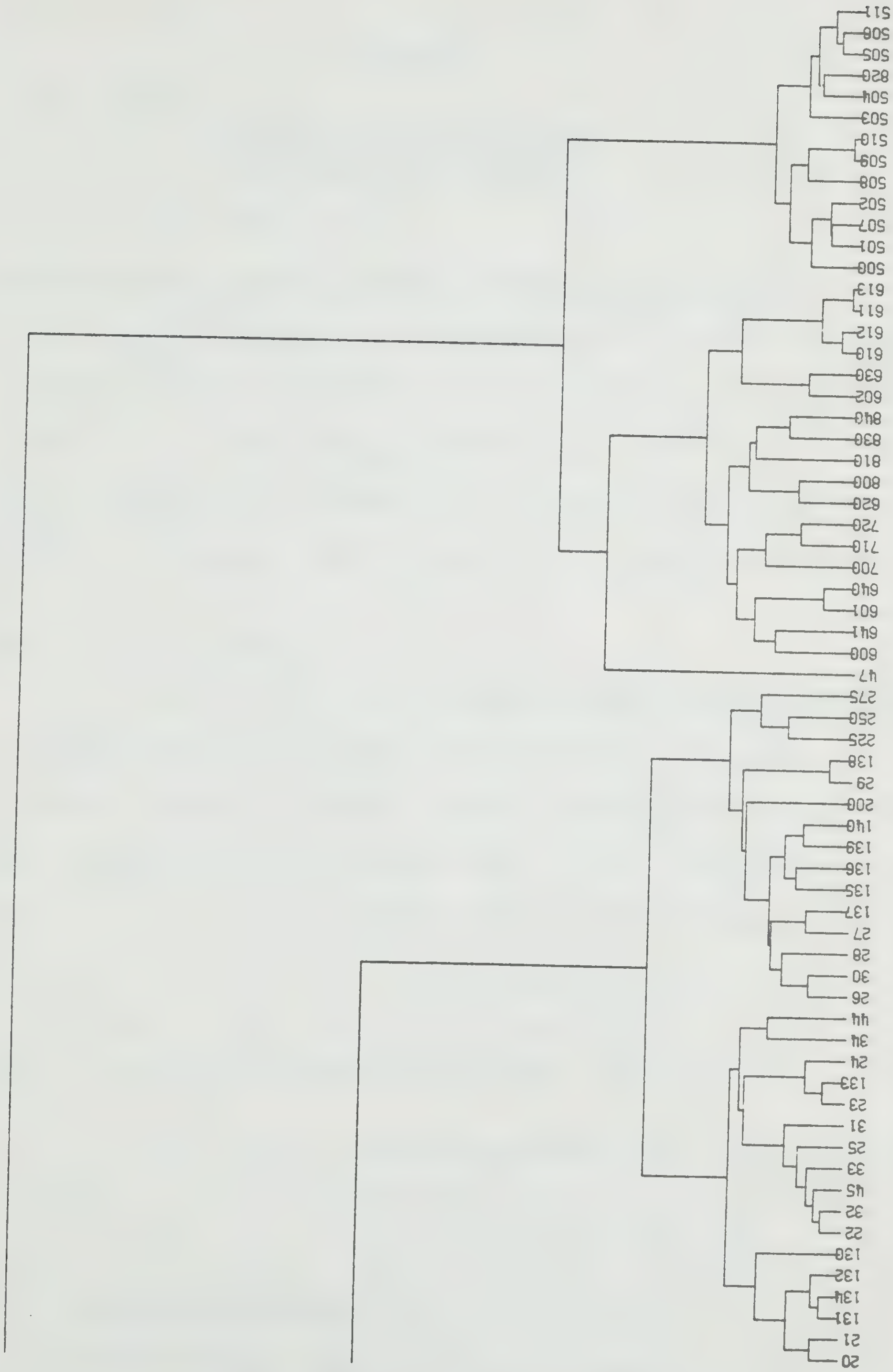

51
45
43
42
34
41
40
33
22
21
19
16
44
20
9
8
7
6
23
13
24
11
15
5
110
18
17
12
10
4
31
28
37
29
47
39
32
25
38
120
27
36
30
26
14
2
220
210
230
200
35
3
130
50
48
49
46
100
1
LONDON
8.553
7.698
6.842
5.987
5.132
4.276
3.421
2.566
1.711
0.855


ST. CATHARINES
22.564
20.307
18.051
15.795
13.538
11.282
9.026
6.769
4.513
2.256
310
311
309
306
305
308
307
304
303
302
301
330
320
300
220
202
215
213
208
211
210
214
212
209
207
204
203
206
205
201
200
110
16
11
20
4
19
18
6
14
13
9
7
17
12
10
5
8
3
2
15
100
1


HAMILTON
21.057
18.951
16.846
14.740
12.634
10.529
8.423
6.317
4.211
2.106
16
48
66
64
63
65
47
46
41
130
45
44
43
39
38
34
17
42
40
19
12
110
100
49
7
24
6
21
15
22
8
23
5
14
10
120
20
18
9
13
4
3
11
2
1


222
221
220
219
217
216
214
225
224
215
210
213
211
212
206
209
208
207
205
204
203
202
201
200
150
160
140
67
68
36
35
62
52
51
50
33
60
32
69
54
53
31
61
30
29
27
300
90
80
70
59
58
73
57
56
55
28
72
71
26
37
25


50
55
110
54
57
56
53
58
52
51
14
122
62
60
102
101
104
103
100
59
300
124
121
120
13
10
61
12
11
9
8
15
7
46
43
42
41
35
19
48
38
49
37
36
39
40
18
123
3
2
17
16
6
5
4
1
OTTAWA
20.271
18.244
16.217
14.190
12.163
10.135
8.108
6.081
4.054
2.027



CHAPTER 3

GRAVITY MODEL CALIBRATION AND GOODNESS OF FIT STATISTICS

The calibration procedure used to estimate the parameters of alternative forms of the gravity model is described in this chapter. In addition a variety of goodness of fit statistics which may be used to assess the qualities of any calibrated gravity model are described in the second part of the chapter. The relative effectiveness of these alternative statistics are also examined for a range of Ontario census areas.

3.1 Basic Gravity Model Structure

The trip distribution model which is calibrated in subsequent chapters to the 1971 journey to work data for all census areas is a doubly constrained version of the gravity model with a negative exponential deterrence function of the following form:

$$T^*_{ij} = A_i B_j O_i D_j \exp(-\beta_s c_{ij}) \tag{3}$$

where T^*_{ij} = the model estimated trip interchanges between zones i and j

O_i = trip productions of zone i

D_j = trip attractions of zone j

β_s = travel deterrence function parameter specific to all the origin zones in a given calibration sub-region s

c_{ij} = the travel "costs" (distance, time or some combination) for a trip between zones i and j

and

$$A_i = [\sum_j B_j D_j \exp(-\beta_s c_{ij})]^{-1} \quad (4)$$

$$B_j = [\sum_i A_i O_i \exp(-\beta_s c_{ij})]^{-1} \quad (5)$$

two balancing factors which ensure that the trip end constraints

$$O_i = \sum_j T_{ij} \quad (6)$$

and

$$D_j = \sum_i T_{ij} \quad (7)$$

are satisfied.

3.2 Model Calibration Technique

The magnitudes of the balancing factors A_i and B_j are estimated by assuming initially that all B_j are equal to 1 and calculating all of the A_i for a specific set of β_s values. These A_i magnitudes are then substituted into equation (5) to calculate a new set of B_j magnitudes. This process continues in an iterative manner until the product magnitude $A_i B_j$ converges for all i=j pairs. Convergence is assumed when the difference between the products of $A_i B_j$ for two successive iterations is less than 1.0×10^{-7}.

The balancing factors A_i and B_j may be set to 1.0 in order to produce, respectively, the production-constrained and attraction-constrained versions of the gravity model. These forms of the gravity model are calibrated for selected census areas.

The gravity model is calibrated by establishing the magnitudes of β_s which minimize the sum of the absolute differences between the observed and model-estimated ordinates of the trip length frequency distributions.

These frequency distributions are specified in 1 km class intervals and where more than one calibration sub-region is used, the observed and simulated frequency distributions are calculated separately for each β_s using only trips originating from those zones to which each parameter is specific.

The model parameters are estimated using a golden section search technique [5]. With this procedure an initial estimate is made of the maximum and minimum limits for the parameter, β_{01} and β_{02}. Two new values are then chosen such that:

$$\beta_{11} = 0.382(\beta_{02} - \beta_{01}) + \beta_{01} \tag{8}$$

$$\beta_{12} = 0.618(\beta_{02} - \beta_{01}) + \beta_{02} \tag{9}$$

The magnitudes 0.382 and 0.618 are derived from $(\Gamma-1)\Gamma$ and $1/\Gamma$ where $\Gamma = 1.618$ the asymptotic interval reduction factor of the Fibonacci search method. The model estimated trip length frequency distributions are calculated using β_{11} and β_{12} and which ever magnitude gives the worst fit is chosen as the new limit of search on that side of the optimum. For example, if β_{12} were the worst the next two parameter magnitudes would be:

$$\beta_{21} = 0.382(\beta_{12} - \beta_{01}) + \beta_{01}$$

$$\beta_{22} = 0.618(\beta_{12} - \beta_{01}) + \beta_{12}$$

However, because of the properties of the Fibonacci numbers $\beta_{11} = \beta_{22}$ and only one new parameter and therefore one new frequency distribution need be calculated during each interation.

Figure 46 illustrates the steps followed in this search procedure and convergence is obtained when the difference between two consecutive values of β is less than some arbitrarily set number and whichever of

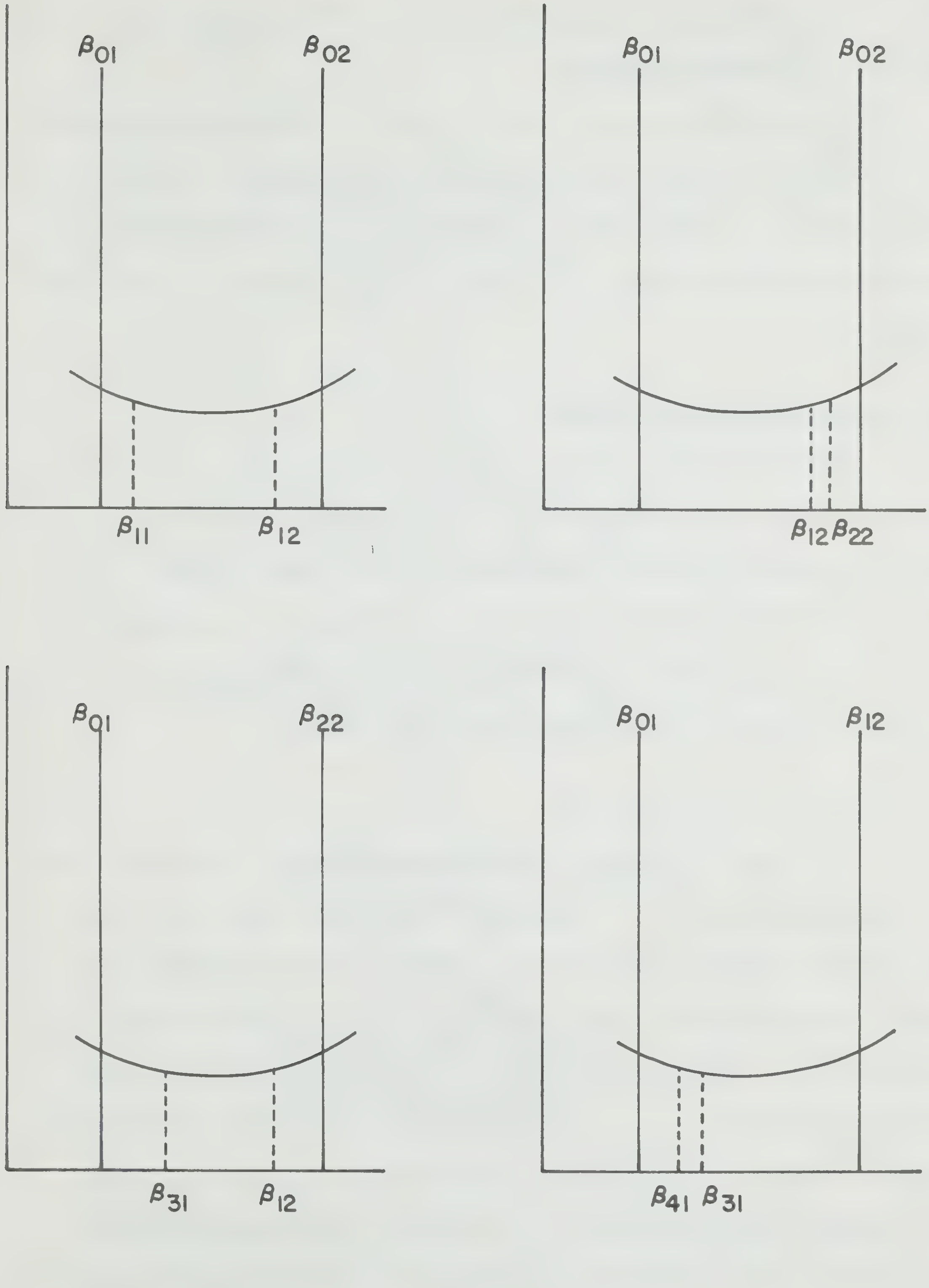


FIGURE 46. Golden Section Search Procedure

these β magnitudes is best according to the calibration procedure is taken as the parameter magnitude.

The adjustment process for A_i and B_j and the estimation of β_s are conducted in an inter-dependent way. The A_i and B_j are adjusted for each β_s before the calculation of the model estimated trip length frequency distribution. The logic of this process is illustrated in Figure 47.

3.3 Alternative Model Statistics

The calibration criterion used in the estimation of the gravity model parameters is the minimization of the differences between the observed and model estimated trip length frequency distributions. However, the observed and model estimated trip matrices may be compared in many ways and the following sections identify some alternative statistics calculated for each of the gravity models calibrated.

3.3.1 Mean Trip Length Statistics

Three mean trip length statistics are reported for each of the calibrated models described in subsequent chapters of this report and these are the observed, simulated and expected mean trip lengths. The observed and simulated mean trip lengths require no explanation and the expected mean trip length is calculated from the trip matrix estimated by equation (3) where $\beta = 0$. It is the mean trip length that would be expected to result from the spatial structure of a community if there were no travel deterrence effect. The expected trip length should be much higher than the simulated or observed where the difference magnitude would be related to the influence of the β magnitude.

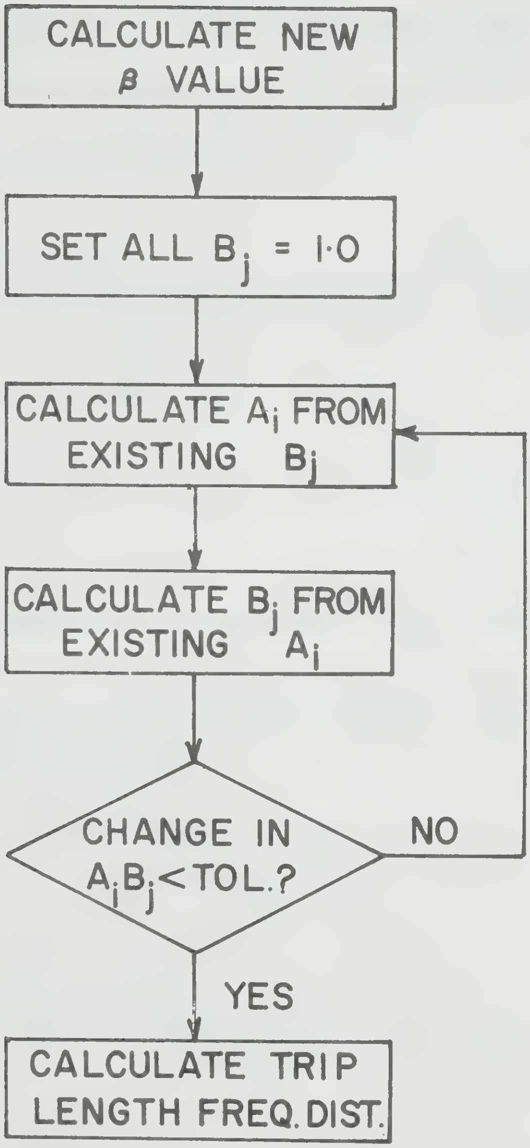


FIGURE 47. Calibration Procedure for Doubly Constrained Gravity Model

Mean trip length statistics are also calculated for each census tract with one set relating to the trips originating from a zone and a second set relating to the trips terminating in a zone. The census tract-specific statistics allow census tracts with inadequate performance to be identified as well as assisting in the assessment of the effectiveness of both trip end balancing and multiple parameter modifications.

3.3.2 Chi-Squared Statistic

The chi-squared statistic is calculated from the model output in the following way:

$$\chi^2 = \sum_i \sum_j \frac{(T_{ij} - T^*_{ij})^2}{T^*_{ij}} \qquad (10)$$

where T_{ij} = the observed trip interchange magnitude between zones i and j

The chi-squared statistic is also calculated for the individual origin and destination zones.

$$\text{Origin-Specific} \quad \chi^2 = \sum_i \frac{(T_{ij} - T^*_{ij})^2}{T^*_{ij}} \qquad (11)$$

$$\text{Destination-Specific} \quad \chi^2 = \sum_j \frac{(T_{ij} - T^*_{ij})^2}{T^*_{ij}} \qquad (12)$$

This is a very sensitive statistic for use in comparing the simulated and observed trip matrices but there are some difficulties with its use. Interchange magnitudes of less than five in either matrix violate the assumptions underlying the chi-squared statistic. In addition the generally poor fit of the model in a statistical sense prevents this

statistic from being used in conventional hypothesis testing.

3.3.3 Phi-Statistic

The phi-statistic is used in information theory and is calculated from

$$\phi = \sum_i \sum_j T_{ij} \left| \log_e \frac{T^*_{ij}}{T_{ij}} \right| \tag{13}$$

The larger the absolute value of phi the poorer is the degree of fit of the model. Since each term in the calculation is multiplied by the observed T_{ij} magnitude this statistic is most sensitive to errors in large interchanges. This statistic may also be calculated for the individual origin and destination zones.

3.3.4 R-Squared Statistics

Two R-squared statistics may be calculated in the following ways:

$$R^2 = 1 - \frac{\sum_i \sum_j (T_{ij} - T^*_{ij})^2}{\sum_i \sum_j (T_{ij} - \bar{T}_{ij})^2} \tag{14}$$

and

$$R^2_{adj} = 1 - \frac{\sum_i \sum_j (T_{ij} - T^*_{ij})^2}{\sum_i \sum_j (T_{ij} - \bar{\bar{T}}_{ij})^2} \tag{15}$$

where $\bar{\bar{T}}_{ij} = \frac{O_i D_j}{T}$

= the expected number of trips between a zonal pair if there were no effect of travel deterrence

Equation (14) compares the differences between the observed and estimated trip interchange residuals with the differences between the observed

and mean trip interchange magnitudes. In this equation R^2 measures the effects of O_i, D_j, A_i, B_j and β on the degree to which T^*_{ij} is an improvement on $\bar{T}_{ij}$. On the other hand equation (15) tends to measure the direct effects of β, or the degree to which T^*_{ij} is an improvement on $\bar{\bar{T}}_{ij}$. Because of this the R^2_{adj} magnitudes will tend to be lower than the R^2 magnitudes.

3.3.5 Likelihood Statistics

Three likelihood statistics may be calculated and these are:

$$L_{obs} = \sum_i \sum_j T_{ij} \log_e \frac{T_{ij}}{T} \tag{16}$$

$$L_{\beta=0} = \sum_i \sum_j T_{ij} \log_e \frac{\bar{\bar{T}}_{ij}}{T} \tag{17}$$

$$L_{sim} = \sum_i \sum_j T_{ij} \log_e \frac{T^*_{ij}}{T} \tag{18}$$

Equation (16) calculates the highest likelihood magnitude that can exist for a given urban area since it operates on the observed T_{ij} magnitudes. Equation (17) calculates the magnitude of the likelihood that would be given by a trip distribution model without interaction deterrence effects. Finally, equation (18) calculates the likelihood for a model with optimum β and this likelihood magnitude should fall between the two extremes.

3.3.6 Other Error Statistics

A variety of other error statistics may be calculated and these include the following traditional measures.

Mean Error

$$MERR = \sum_i \sum_j (T_{ij} - T^*_{ij})/n^2 \tag{19}$$

Mean Absolute Percentage Error

$$MABSPC = \sum_i \sum_j \left| \frac{T_{ij} - T^*_{ij}}{T_{ij}} \right| .100/n^2 \qquad (20)$$

Standard Deviation of Residuals

$$SDRES = [\sum_i \sum_j \{\frac{(T_{ij} - T^*_{ij}) - MERR}{n^2 - 1}\}^2]^{1/2} \qquad (21)$$

Total Absolute Error

$$TABSERR = \sum_i \sum_j |T_{ij} - T^*_{ij}| \qquad (22)$$

Mean Absolute Error

$$MABSERR = TABSERR/n^2 \qquad (23)$$

3.4 Detailed Analyses of Chi-Squared and Phi Statistics

Previous analyses have shown that the chi-squared and phi statistics are very sensitive when used in the trip table context and large magnitudes are obtained. In order to examine further the behaviour of chi-squared and phi statistics Table 6 was generated in which the rows represent observed magnitudes of T_{ij} from 25 to 500 trips in increments of 25 trips and the columns represent simulated T^*_{ij} magnitudes over the same range. The entries represent the contribution to chi-squared for each T_{ij} versus T^*_{ij} comparison. The entries above the diagonal represent over-estimation residuals while the entries below the diagonal represent under-estimation residuals. Inspection of the entries in the matrix demonstrates clearly the asymmetric character of chi-square with the under-estimation errors being more important

TABLE 6. Single Interchange Values of Chi-Square

T_{ij} T^*_{ij}	25	50	75	100	125	150	175	200	225	250	275	300	325	350	375	400	425	450	475	500
25	0	12	33	56	80	104	128	153	178	202	227	252	277	302	327	351	376	401	426	451
50	25	0	8	25	45	67	89	112	136	160	184	208	233	257	282	306	331	355	380	405
75	100	12	0	6	20	37	57	78	100	122	145	169	192	216	240	264	288	312	337	361
100	225	50	8	0	5	17	32	50	69	90	111	133	156	178	202	225	248	272	296	320
125	400	112	33	6	0	4	14	28	44	62	82	102	123	145	167	189	212	235	258	281
150	625	200	75	25	5	0	3	12	25	40	57	75	94	114	135	156	178	200	222	245
175	900	312	133	56	20	4	0	3	11	22	36	52	69	87	107	126	147	168	189	211
200	1225	450	208	100	45	17	3	0	3	10	20	33	48	64	82	100	119	139	159	180
225	1600	612	300	156	80	37	14	3	0	2	9	19	31	45	60	76	94	112	131	151
250	2025	800	408	225	125	67	32	12	3	0	2	8	17	28	42	56	72	89	106	125
275	2500	1012	533	306	180	104	57	28	11	2	0	2	8	16	27	39	53	68	84	101
300	3025	1250	675	400	245	150	89	50	25	10	2	0	2	7	15	25	37	50	64	80
325	3600	1512	833	506	320	204	128	78	44	22	9	2	0	2	7	14	23	35	47	61
350	4225	1800	1008	625	405	267	175	112	69	40	20	8	2	0	2	6	13	22	33	45
375	4900	2112	1200	756	500	337	228	153	100	62	36	19	8	2	0	1	6	12	21	31
400	5625	2450	1408	900	605	417	289	200	136	90	57	33	17	7	2	0	1	5	12	20
425	6400	2812	1633	1056	720	504	357	253	178	122	82	52	31	16	7	1	0	1	5	11
450	7225	3200	1875	1225	845	600	432	312	225	160	111	75	48	28	15	6	1	0	1	5
475	8100	3612	2133	1406	980	704	514	378	278	202	145	102	69	45	27	14	6	1	0	1
500	9025	4050	2408	1600	1125	817	603	450	336	250	184	133	94	64	42	25	13	5	1	0

than the over-estimation errors.

Figure 48 shows the variation in the chi-square magnitude with T_{ij} magnitude for constant error levels of ±50, ±100, ±200. The diagram illustrates very clearly the rapid increase in the chi-square magnitude as the under-estimation error approaches the value of T_{ij}, that is as the value of T_{ij}^* approaches zero. This indicates the high sensitivity of chi-square to under-estimation errors when T_{ij}^* is small which is a frequent occurrence in most simulated trip matrices.

Table 7 shows the phi statistic magnitudes for different combinations of observed and simulated trip interchange magnitudes. Inspection of Table 7 shows that the phi statistic is also asymmetric but less extreme than the chi-squared statistic. Figure 49 shows the variation in phi with the observed trip interchange magnitude for different error magnitudes. Although phi has a greater sensitivity to under-estimation than over-estimation this is not nearly as severe as was evident in the analysis of chi-squared. It should be noted also that the rapid increase of the statistic when T_{ij}^* approaches zero is most marked for large values of T_{ij}. The phi statistic is insensitive to errors involving small interchange values. Inspection of the over-estimation values reveals this property even more clearly. The phi values increase as T_{ij} increases despite the fact that the relative size of the error as compared to the observed value decreases. This is a very useful property in a statistic being used to evaluate trip distribution models. In most cases the observed matrix has been synthesized using a limited sample which often results in low levels or reliability for small interchange values. The phi statistic automatically accounts for some of this unreliability by discounting the effect of errors in small interchanges.

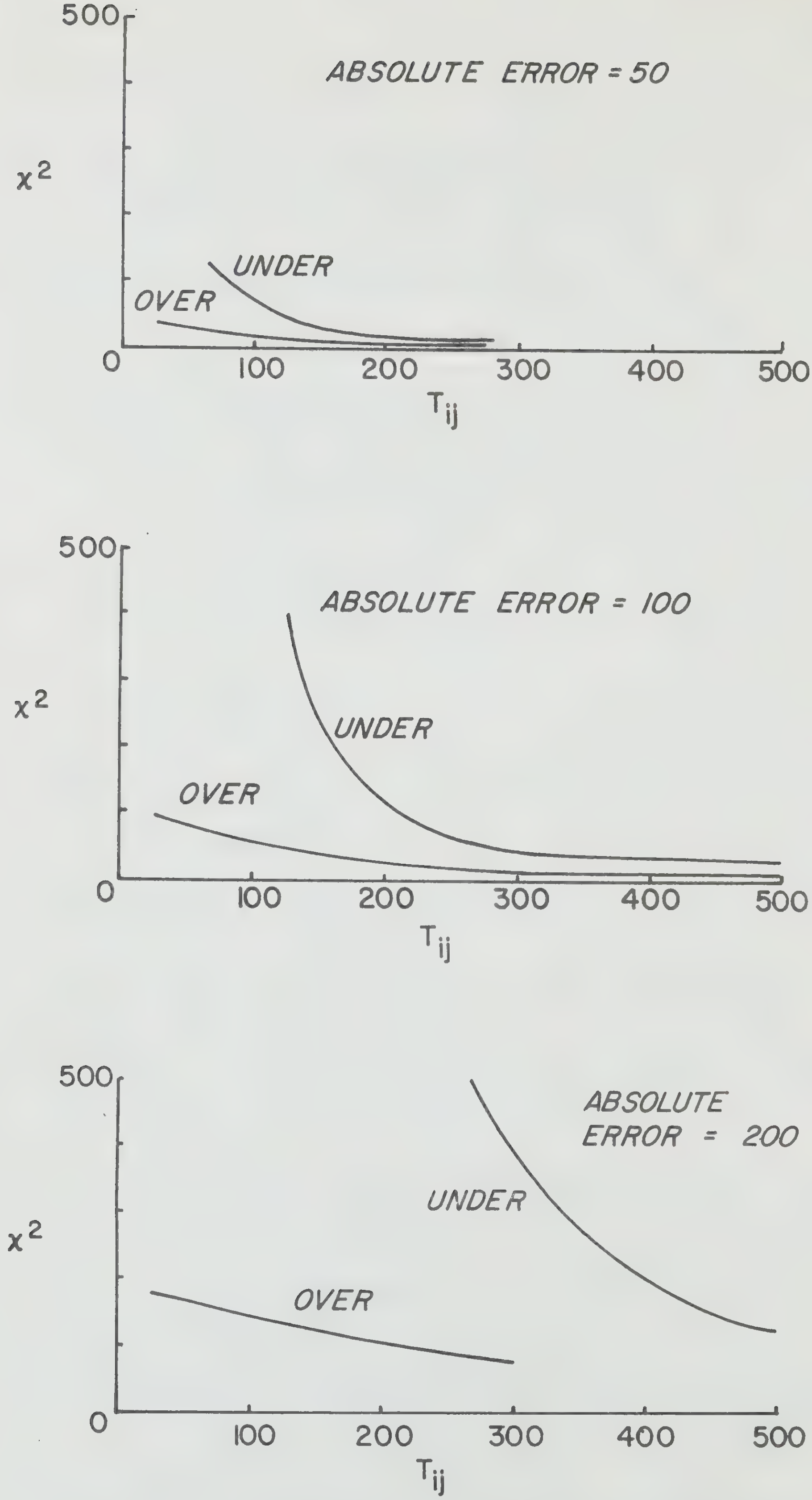


FIGURE 48. Variation in Chi-Squared Magnitude with Observed Trip Interchange Magnitude

TABLE 7. Single Interchange Values of Phi

T_{ij} \ T^*_{ij}	25	50	75	100	125	150	175	200	225	250	275	300	325	350	375	400	425	450	475	500
25	0	17	27	35	40	45	49	52	55	57	60	62	64	66	68	69	71	72	74	75
50	35	0	20	35	46	55	63	69	75	80	85	89	93	97	101	104	107	110	112	115
75	82	30	0	21	38	52	63	73	82	90	97	104	110	115	121	125	130	134	138	142
100	139	69	29	0	22	40	56	69	81	92	101	110	118	125	132	139	145	150	156	161
125	201	114	84	28	0	23	42	59	73	87	98	109	119	129	137	145	153	160	167	173
150	269	165	104	61	27	0	23	43	61	77	91	104	116	127	137	147	156	165	173	180
175	340	219	148	98	59	27	0	23	44	62	79	94	108	121	133	145	155	165	175	184
200	416	277	196	139	94	57	27	0	23	45	64	81	97	112	126	139	151	162	173	183
225	494	338	247	182	132	91	56	26	0	24	45	65	83	99	115	129	143	156	168	180
250	576	402	301	229	173	128	89	56	26	0	24	45	65	84	101	117	133	147	160	173
275	659	469	357	278	217	167	124	87	55	26	0	24	46	66	85	103	120	135	150	164
300	745	537	416	329	263	208	162	122	86	55	26	0	24	46	67	86	104	122	138	153
325	834	608	476	383	310	251	201	158	119	85	54	26	0	24	46	67	87	106	123	140
350	924	881	539	438	360	296	243	196	155	118	84	54	26	0	24	47	68	88	107	125
375	1015	755	603	496	412	344	286	236	191	152	116	84	54	26	0	24	47	68	89	108
400	1109	832	669	554	465	392	331	277	230	188	150	115	83	53	26	0	24	47	69	89
425	1204	909	737	685	520	443	377	320	270	225	185	148	114	82	53	26	0	24	47	69
450	1301	989	806	677	576	494	425	365	312	264	222	182	146	113	82	53	26	0	24	47
475	1399	1069	877	740	634	547	474	411	355	305	260	218	180	145	112	82	53	26	0	24
500	1498	1151	948	805	693	602	525	458	399	346	299	255	215	178	144	111	81	53	26	0

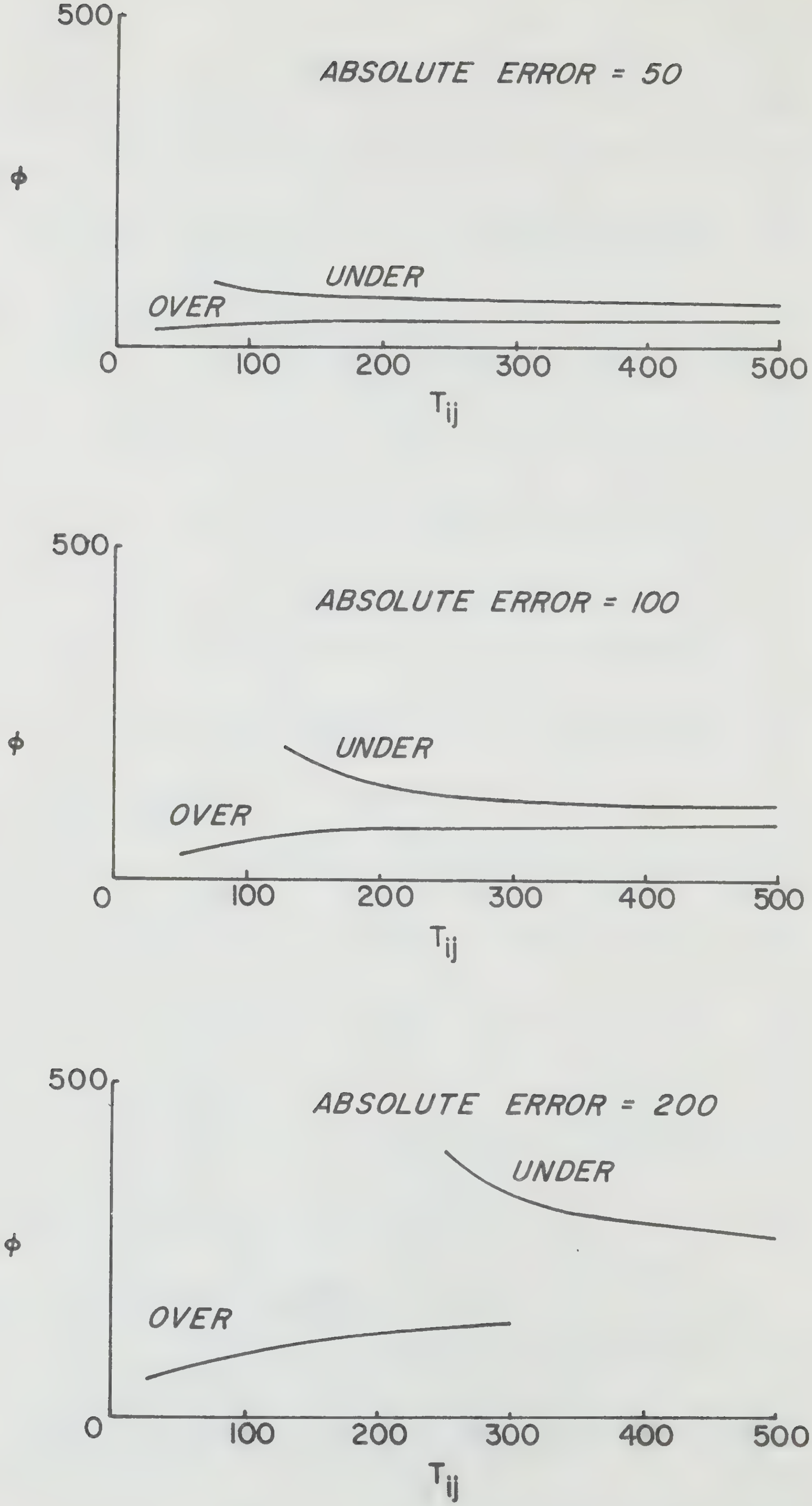


FIGURE 49. Variation in Phi Magnitude with Observed Trip Interchange Magnitude

The analysis of the response characteristics of the chi-squared and the phi statistics at the interchange level indicates that each responds in a slightly different manner when the error changes. Both measures have a greater sensitivity to under-estimation errors than to over-estimation errors with this characteristic being less severe for phi. Phi is relatively more sensitive to over-estimation errors in large interchanges with chi-squared more sensitive to all errors in small interchanges. For under-estimation errors both statistics are more sensitive to an error of a specific magnitude on a smaller interchange.

3.5 Goodness of Fit Statistics for Actual Census Areas

The following sections of this chapter examine the behavioural of the goodness of fit statistics described in Section 3.3 for selected Ontario census areas. The behaviour of the statistics is examined both within individual census areas and between census areas. It is not possible to approach this problem of the identification of the best goodness of fit statistics in the traditional statistical sense. The magnitudes of the statistics are influenced by the number of zones, the total number of trips and the characteristic probability distributions of the observed and estimated trip interchange magnitudes. A simulation approach has been taken in order to evaluate the performance of the alternative statistics.

3.6 The Simulation Method

Eight Ontario census areas were selected for the detailed evaluation of model statistics and Table 8 lists these areas along with the population, number of census tracts and total number of observed trips. The areas selected represent a range of population sizes and spatial structures.

TABLE 8. Characteristics of the Eight Ontario Census Areas, 1971

Census Area	Population In Official Census Tracts	Number of Census Tracts	Number of Trips	Mean Interchange Magnitude
Oshawa	116,911	22	33,740	70
Thunder Bay	108,411	25	33,130	53
Kitchener	212,819	45	81,095	40
Windsor	211,494	56	74,455	24
London	223,222	59	97,195	30
St. Catharines	221,282	53	91,815	33
Hamilton	437,554	109	152,265	13
Ottawa	506,316	120	201,015	14

Simulated trip interchange matrices were calculated by multiplying the observed trip interchange entries by a random percentage error where these errors were generated from a rectangular distribution with mean of zero and a specified range. Six percentage ranges were used and these were ten, twenty-five, fifty, seventy-five, one hundred and one hundred and fifty percent. The random number generator GGU3 from the IMSL subroutine library was used and a simulated matrix calculated from the following equation:

$$T^*_{ij} = T_{ij} + \delta.T_{ij}.\text{RND}.\text{FACT} \qquad (27)$$

where δ = a random number taking values of +1 or -1

RND = a random number where $0 \leq \text{RND} \leq 1.0$

FACT = allowable percentage error divided by 100

The same starting value or "seed" was used for each set of simulations for each city and the estimated trip interchange matrix was simulated three times.

The use of the same starting numbers made it possible to evaluate each statistic with respect to the variations in the allowable error only. The multiple simulations were used to ensure that these evaluations were robust and were not dependent upon a particular sequence of random numbers that happened to be generated. The simulated trip interchange matrices were then adjusted so that the row and column totals equalled the observed labour force and employment magnitudes. In this sense the simulated trip interchange matrices were similar to those produced by a doubly constrained gravity model.

3.7 Behaviour of Alternative Statistics

Figure 50 shows the variation in the R^2 and R^2_{adj} magnitudes with allowable error percentage for the eight census areas. The tabulated

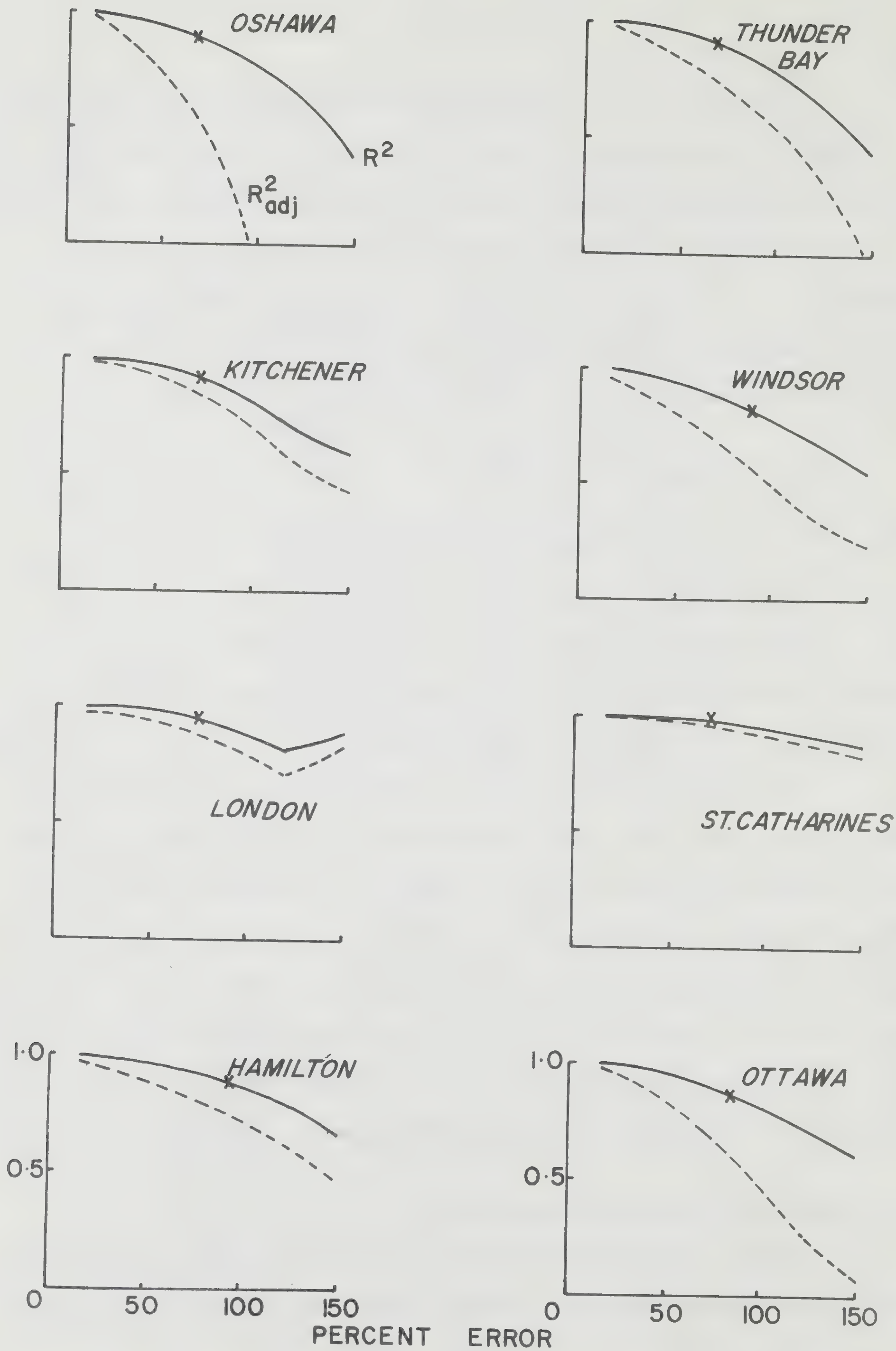


FIGURE 50. Variation in R^2 and R^2_{adj} with Allowable Error Percentage

magnitudes are presented in the appendix of this chapter. This diagram shows that in all cases the R^2_{adj} statistic is lower than R^2 and more sensitive to error increase. It is interesting to note that even when the allowable percentage error is 100 percent the R^2 magnitudes are greater than about 0.7. The diagram also illustrates that R^2 is not a very sensitive statistic for comparing the observed and simulated trip matrices in the larger cities. This is also the case for the R^2_{adj} in the larger cities except for Ottawa where it continues to decrease with increasing error percentage.

The star marked on each plotted R^2 line indicates the R^2 magnitudes obtained for the gravity models calibrated for each of the eight census areas. The locations of these points would suggest estimated trip matrices with errors in the range of 60-70 percent except for Windsor, Hamilton and Ottawa where error magnitudes of around 90 percent are indicated.

Figure 51 shows the variation in the chi-squared statistic with percentage error for the eight census areas and the tabulated values are presented in the appendix. This diagram illustrates that the chi-squared magnitude increases sharply beyond an allowable percentage error magnitude of 75 percent. Beyond an error magnitude of 100 percent the increase in chi-squared tends to level off and this reflects the fact that negative trip interchange estimates are not allowed and the under-estimations are constrained to a maximum of 100 percent. The stars shown on the diagram identify the chi-squared magnitudes for the calibrated models and these correspond generally to error magnitudes of 75-100 percent.

Figure 52 shows the variation in the magnitude of the phi statistic with increasing percentage error. It is important to note that up to error magnitudes of 100 percent the phi statistic increases almost linearly with

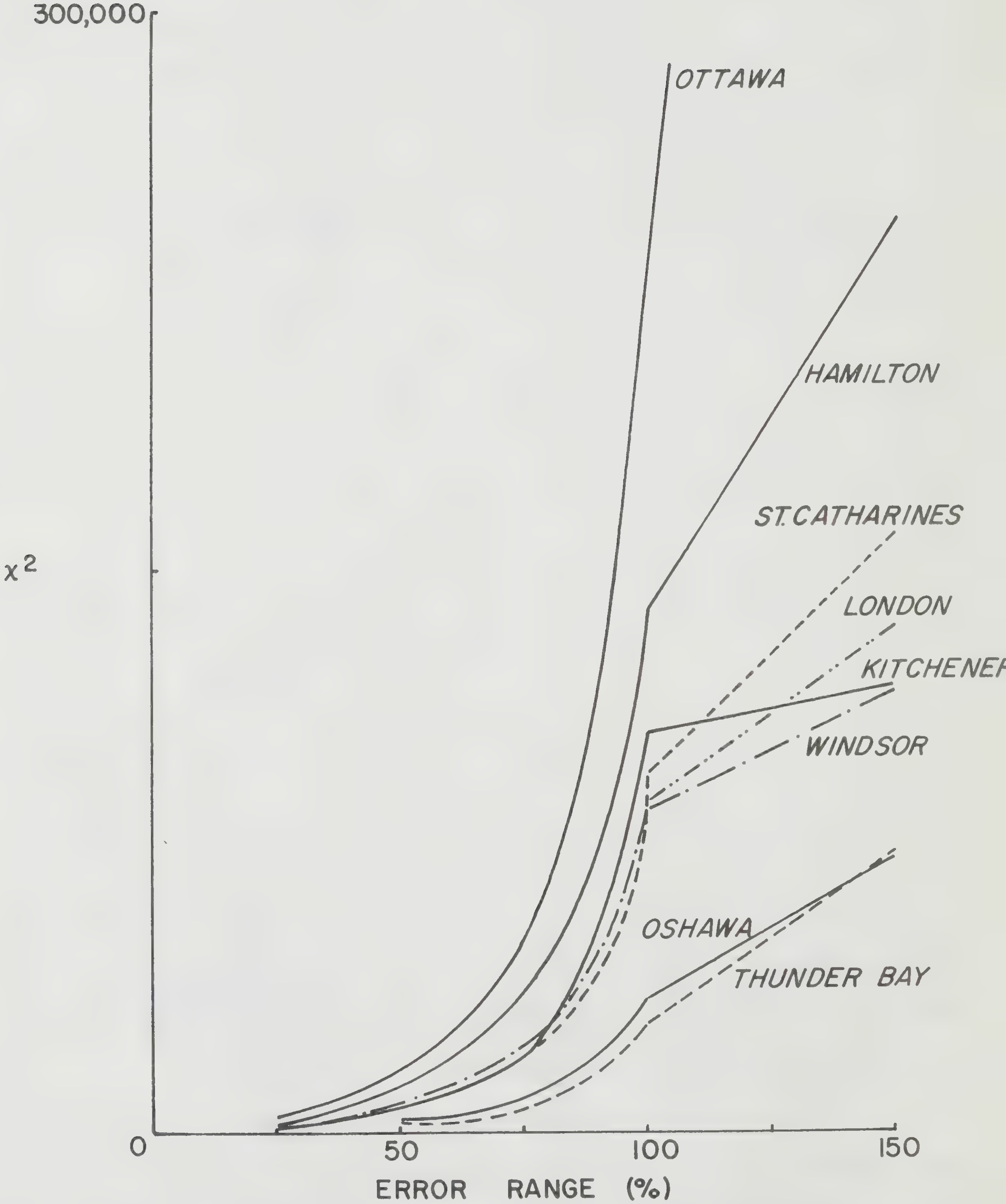


FIGURE 51. Variation in Chi-Squared with Allowable Error Percentage

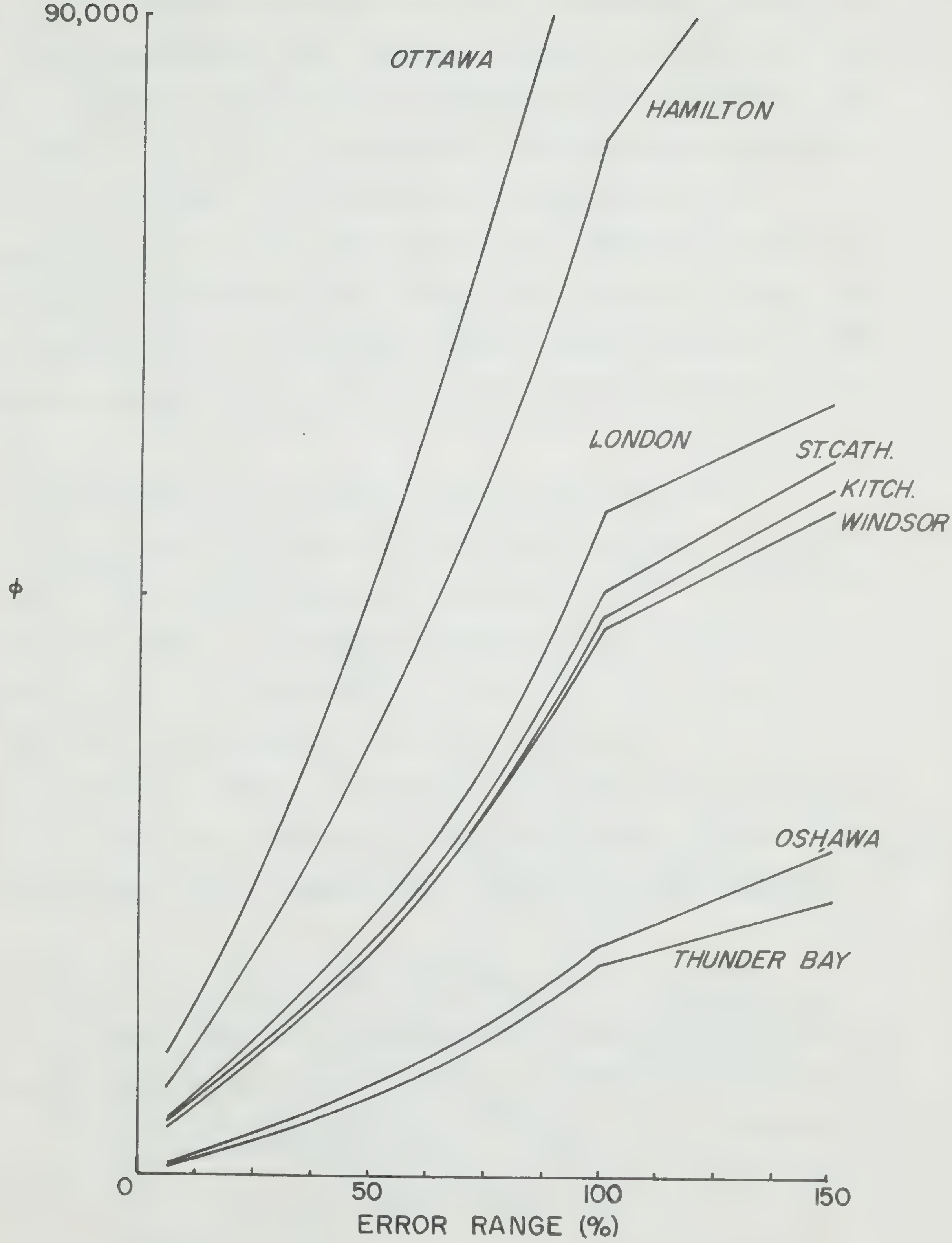


FIGURE 52. Variation in Phi with Allowable Error Percentage

increasing percentage error. The phi statistic also appears to be more stable than chi-squared in that the ranking of cities is consistent across all error magnitudes except for Kitchener and Windsor. The phi magnitudes for the doubly constrained models estimated for each census area are shown on the diagram suggesting interchange error magnitudes of the order of 80-90 percent.

Figure 53 illustrates the variation in the likelihood statistics with increasing percentage error for the eight census areas and the tabulated magnitudes are presented in the appendix. The relationships illustrated show that the likelihood statistic behaves very similar to the chi-squared statistic with the magnitude increasing very sharply beyond error magnitude percentages of about 75 percent.

Figure 54 shows the standard deviation of the residuals with increasing error magnitude. This diagram illustrates that the increase is roughly linear with the exception of London where the magnitude decreases with a one hundred and fifty percent allowable error. A similar behaviour may be observed in the R^2 statistic for London and this behaviour occurs only in the first of the three simulations and appears to be a result of the particular random number set generated. It is also interesting to note that the standard deviation decreases with increasing census area size and this is because the average size of an interchange is much larger for the smaller cities.

Figure 55 shows the variation in the mean absolute error of a trip interchange magnitude with increasing trip interchange error. The mean absolute error behaves in a very similar way to the standard deviation of the errors.

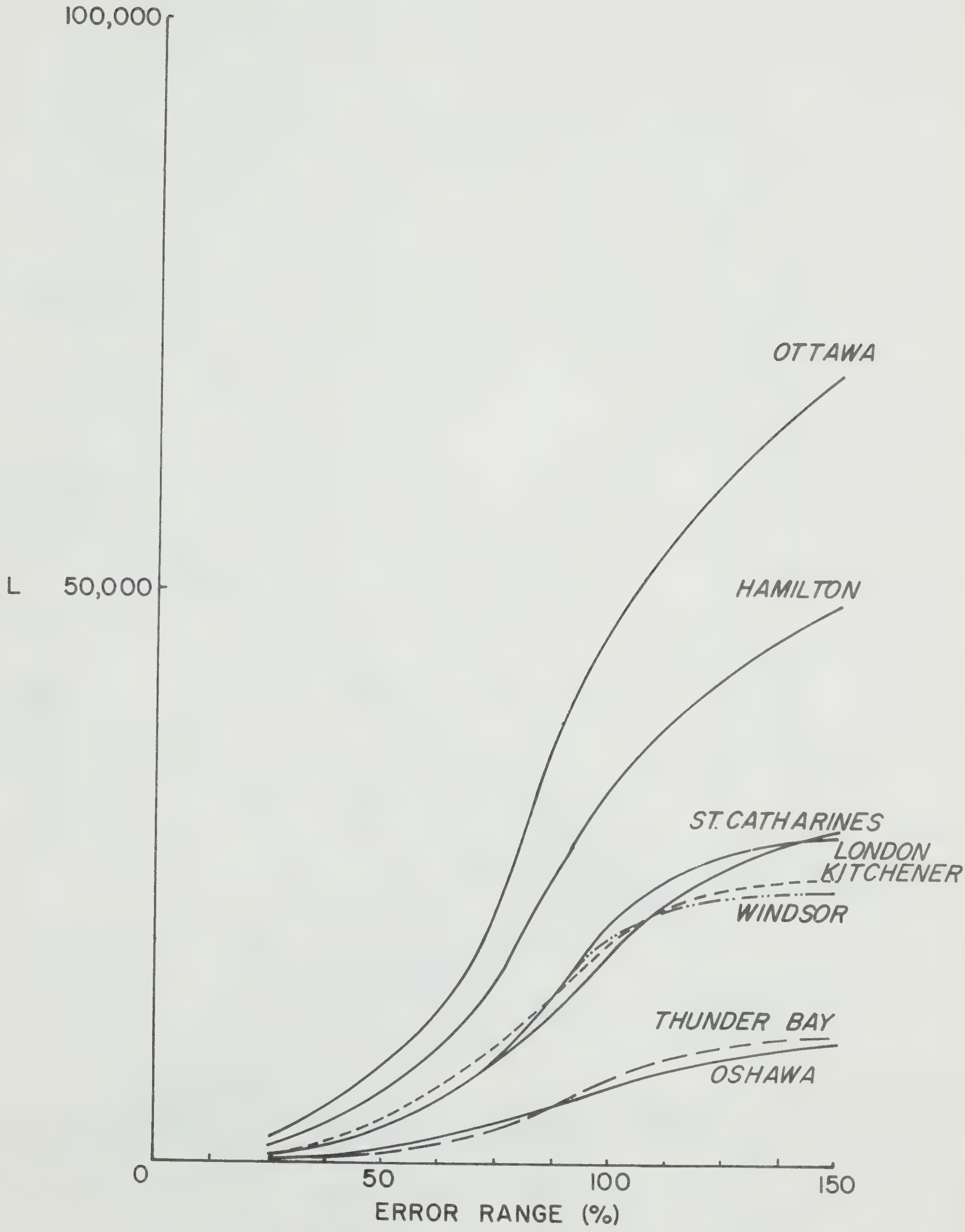


FIGURE 53. Variation in Likelihood with Allowable Error Percentage

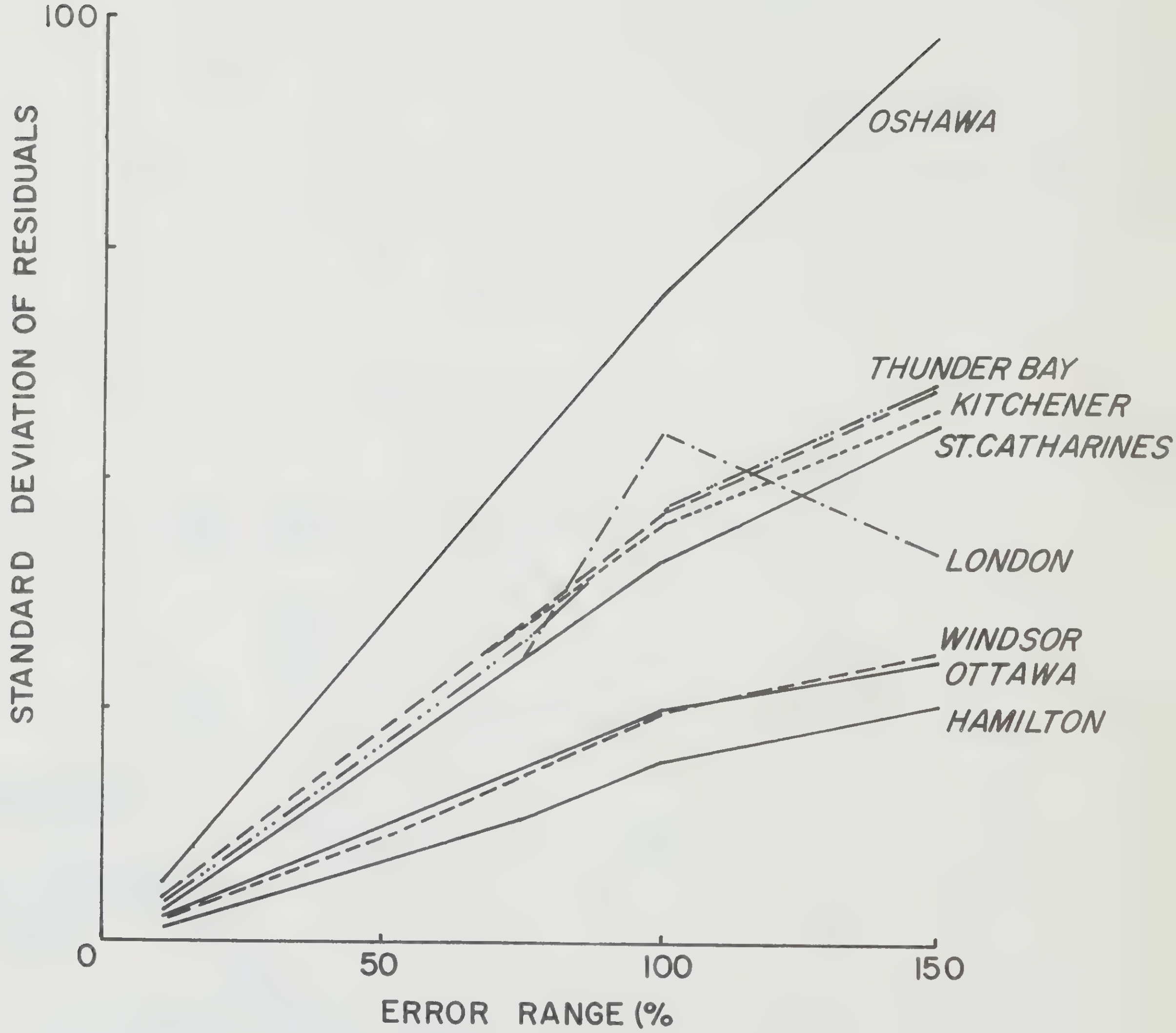


FIGURE 54. Variation in Standard Deviation of Residuals with Allowable Error Percentage

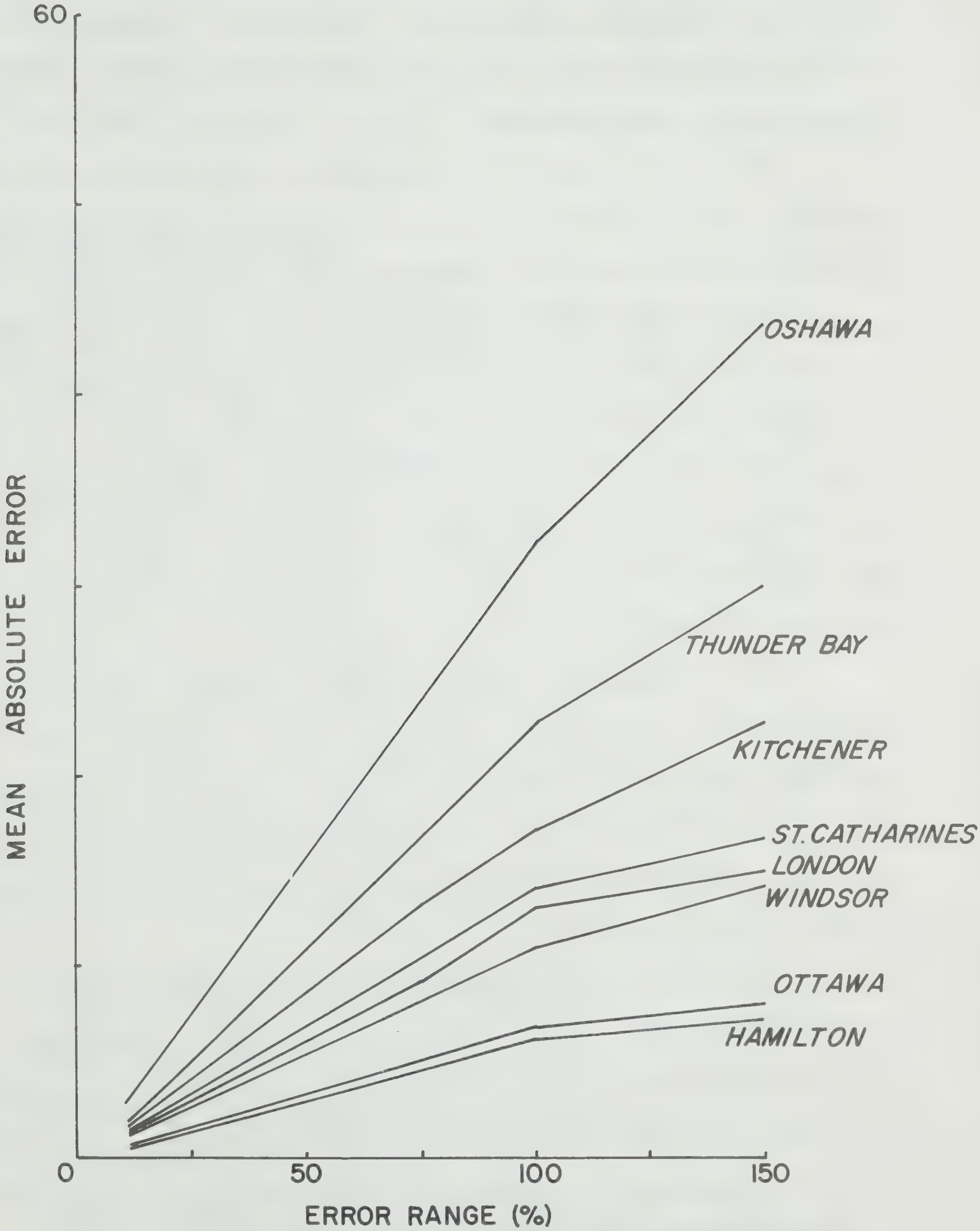


FIGURE 55. Variation in Mean Absolute Error with Allowable Error Percentage

3.8 Comparative Behaviour of Statistics

Three criteria may be identified for evaluating the quality of the alternative statistics and these are: (i) sensitivity to changes in error; (ii) similarity in magnitude across cities with different sizes; and (iii) consistency with other statistics.

The graphs of each of the statistics as a function of the maximum percentage error have shown that phi, the standard deviation of the residuals and the mean absolute error are equally sensitive to change in the error across all magnitudes of the error percentage. The chi-squared and likelihood statistics increase exponentially with increasing error magnitude and do not provide good discrimation at low error magnitudes. The R^2 statistic had very poor sensitivity to error magnitude changes remaining at high levels for large changes in the error magnitudes. The R^2_{adj} statistic is a more sensitive statistic for some of the urban areas but was insensitive in some urban areas particularly London and St. Catharines.

With the exception of the R^2 and R^2_{adj} statistics none of the statistics may be used to compare the goodness of fit of models across all urban areas since the magnitudes of the statistics are sensitive to urban area size, or the number of trip interchanges entries being compared. However, some transformations of the statistics are suggested in the following sections to allow comparisons across urban areas.

The third criterion suggested above is consistency. The chi-squared statistic appears to be inconsistent for large error magnitudes since the error functions cross-over at large percentage error magnitudes. Also in one of the simulations for R^2 and standard deviation of the residual measures are inconsistent for London where both magnitudes decreased when the error percentage increased from 100 to 150 percent.

Table 9 summarizes the ratings of each of the statistics with respect to the three criteria. This would suggest that the phi statistic and the mean absolute error statistics are perhaps the best statistics to use. However, their principal deficiency is that they could be used across urban areas to compare model behaviour.

3.9 Comparisons Across Urban Areas

Chi-squared and phi statistics have been calculated for all of the Ontario census areas as well as for some stratified matrices for Kitchener, Hamilton and Ottawa using a maximum random error range of 75 percent. Figures 56 and 57 show the magnitudes of chi-squared and phi plotted against the number of home to work linkages in each area. These relationships indicate underlying linear trends suggesting that a linear transformation of either measure may be used as an indication of model performance across urban areas. Perhaps the most useful measure, then, is the phi statistic because of its constant sensitivity to error magnitude.

Figure 58 shows R^2 and the phi statistic divided by the number of interchanges plotted against the number of interchanges when the estimated matrix has been generated using a 75 percent error magnitude. This diagram shows that the transformed phi statistic is quite consistent across all urban area sizes while there is a band of variation of R^2.

The behaviour of these two statistics for the calibrated gravity models is shown in Figure 59. This graph shows that the transformed phi statistic increases with urban area size indicating that model performance is poorer with increasing urban area size.

Another possibility would be to use the standard deviation of the errors and the mean absolute error statistics to characterize the goodness

TABLE 9. Summary of Effectiveness of Goodness of Fit Statistics

	R^2	R^2_{adj}	χ^2	Phi	Likelihood	Standard Deviation of Residuals	Mean Absolute Error
Sensitivity to Error Independent of Error	No	No	No	Yes	No	Yes	No
Level of Sensitivity	Low	Medium	Very High	High	High	High	High
Similar Magnitudes For All Cities	Yes	Yes	No	No	No	No	No
Consistent With Other Measures	No	No	No	Yes	Yes	No	Yes

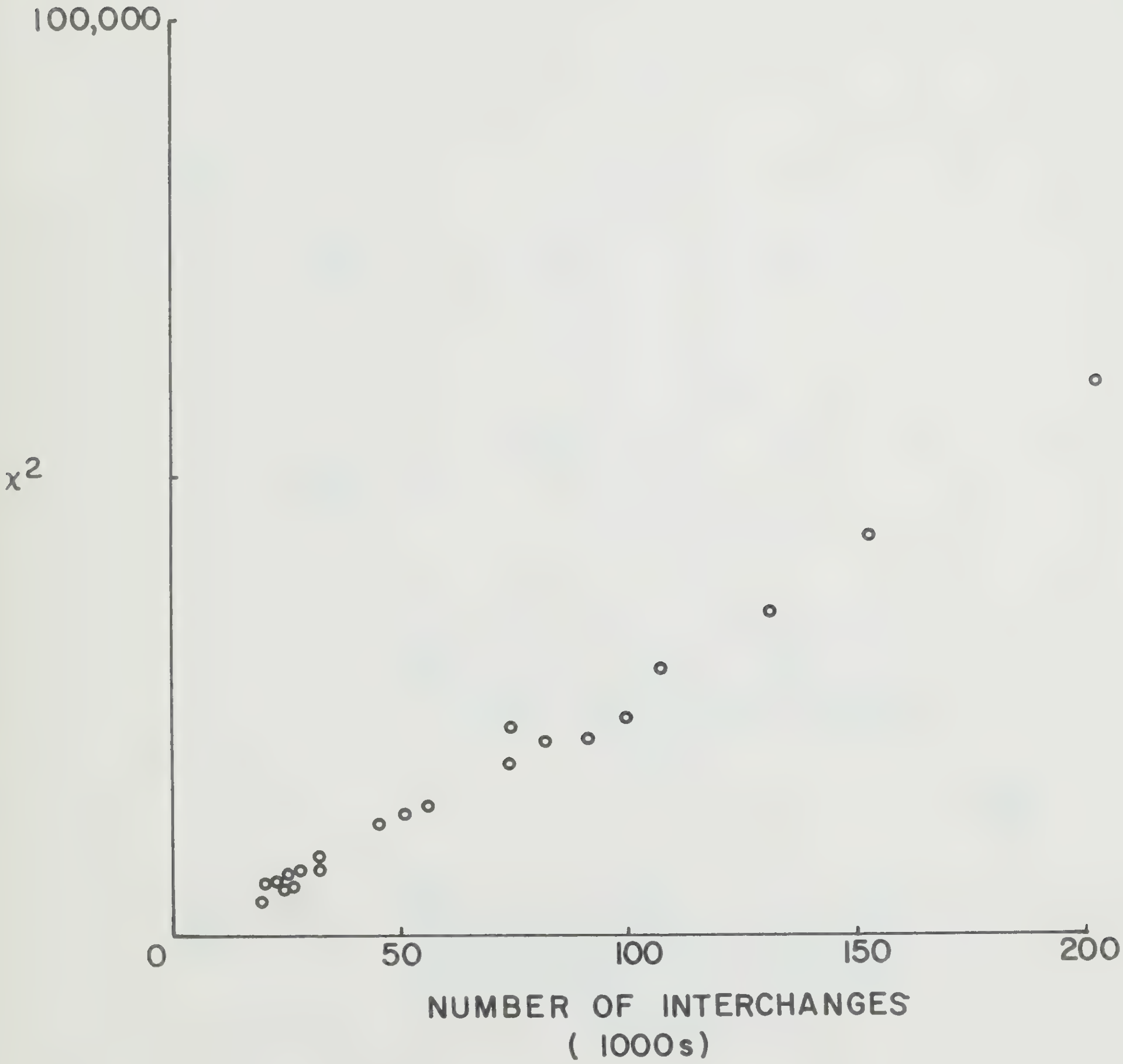


FIGURE 56. Variation in Chi-Squared with Number of Trip Linkages for 75 Percent Error Magnitude

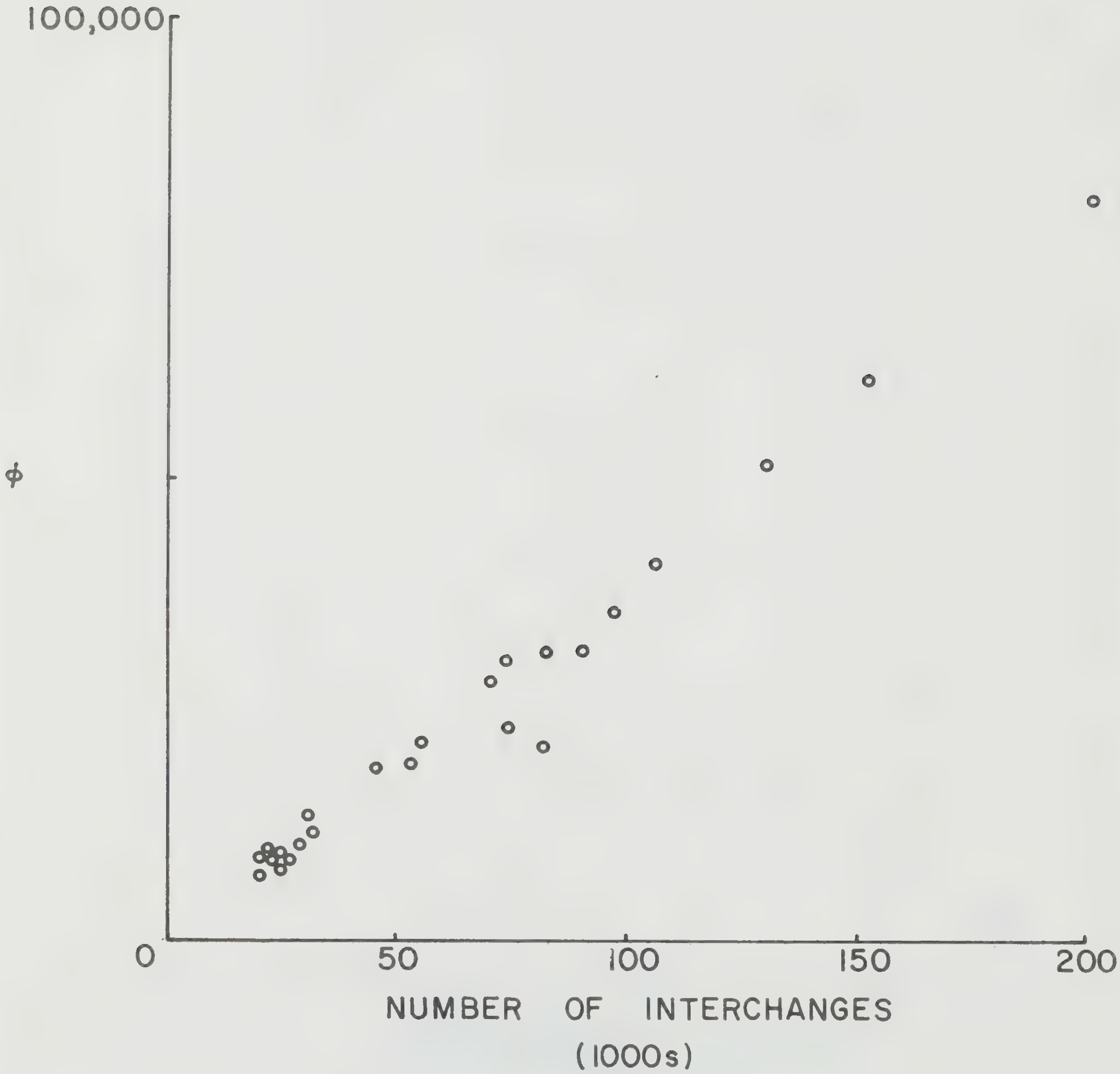


FIGURE 57. Variation in Phi with Number of Trip Linkages for 75 Percent Error Magnitude

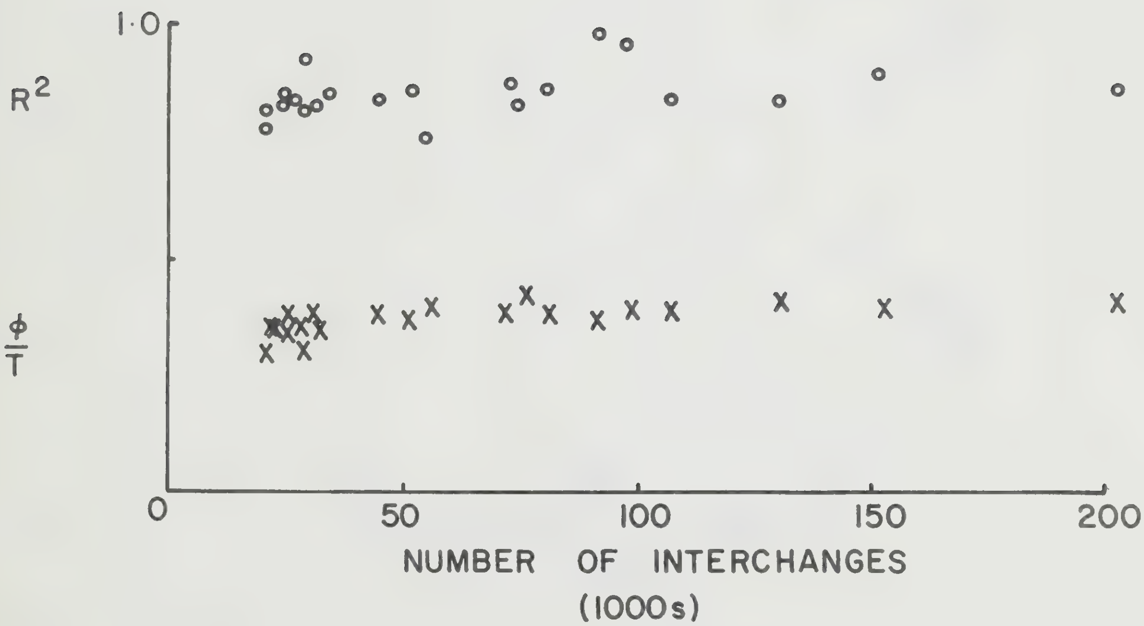


FIGURE 58. Variation in R^2 and Phi/T with Number of Trip Interchanges for a 75 Percent Error

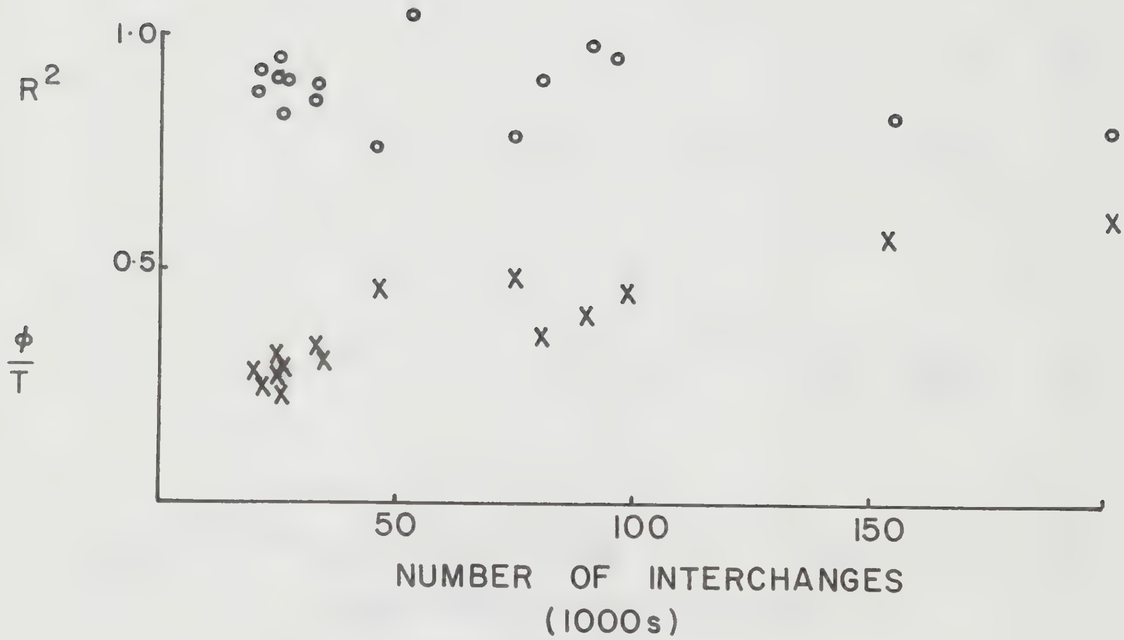


FIGURE 59. Variation in R^2 and Phi/T with Number of Trip Interchanges for Single Parameter Doubly Constrained Gravity Model

of fit of trip distribution models. However, an evaluation of these statistics across urban areas indicated that they had certain undesirable characteristics. The analyses presented above would suggest that the transformed phi statistic is the best goodness of fit measure to use. This statistic is used only as evaluative statistic in this report but its properties would suggest that it should be used as a calibration criterion as well.

APPENDIX TO CHAPTER 3

Error Statistic Summaries

TABLE 3.1. Performance of R^2 and R^2_{adj}

Census Area	10%		25%		50%		75%		100%		150%	
	1.0	.98	.98	.89	.92	.58	.81	.03	.66	-.81	.34	-1.48
Oshawa	1.0	.99	.98	.91	.93	.63	.84	.14	.69	-.64	.53	-1.47
	1.0	.99	.98	.92	.93	.66	.84	.17	.69	-.64	.56	-1.34
	1.0	.99	.98	.96	.93	.85	.83	.65	.66	.30	.44	- .14
Thunder Bay	1.0	.99	.98	.96	.93	.85	.82	.64	.65	.29	.47	- .09
	1.0	.99	.98	.95	.90	.79	.76	.51	.54	.05	.30	- .44
	1.0	1.0	.99	.98	.94	.92	.86	.81	.73	.64	.58	.43
Kitchener	1.0	1.0	.99	.98	.94	.92	.84	.78	.61	.48	.51	.34
	1.0	1.0	.99	.98	.94	.93	.87	.83	.75	.67	.60	.47
	1.0	1.0	.98	.97	.93	.88	.83	.72	.69	.49	.56	.26
Windsor	1.0	.99	.98	.97	.93	.87	.81	.69	.63	.39	.55	.26
	1.0	1.0	.97	.97	.93	.88	.83	.72	.67	.45	.54	.23
	1.0	1.0	1.0	.99	.98	.97	.94	.92	.81	.75	.89	.85
London	1.0	1.0	1.0	1.0	.99	.98	.97	.95	.93	.91	.90	.87
	1.0	1.0	1.0	1.0	.99	.98	.96	.95	.93	.91	.91	.89
	1.0	1.0	1.0	1.0	.98	.98	.96	.96	.93	.92	.88	.86
St. Catharines	1.0	1.0	1.0	1.0	.98	.98	.95	.94	.89	.87	.96	.84
	1.0	1.0	1.0	1.0	.99	.98	.97	.96	.93	.92	.78	.74
	1.0	1.0	.99	.98	.95	.91	.89	.80	.79	.64	.66	.42
Hamilton	1.0	1.0	.99	.98	.94	.90	.86	.76	.72	.53	.65	.41
	1.0	1.0	.99	.98	.95	.91	.88	.79	.77	.60	.71	.50
	1.0	1.0	.99	.97	.94	.87	.86	.63	.73	.36	.62	.11
Ottawa	1.0	1.0	.99	.97	.95	.88	.88	.71	.77	.46	.66	.20
	1.0	.99	.99	.97	.94	.86	.85	.66	.69	.28	.66	.20

Percentage Error (column heading over 10%–150%)

Cell entries are R^2 and R^2_{adj},

TABLE 3.2. Performance of Chi-Squared

Census Area	Percentage Error					
	10%	25%	50%	75%	100%	150%
	105	635	2835	8645	43838	91660
Oshawa	113	630	2804	8220	37447	94628
	101	605	2739	8214	41602	84601
	91	543	2483	7635	35387	91803
Thuncer Bay	105	602	2780	8731	54901	57907
	113	684	3092	9450	45057	99965
	281	1578	7026	21431	128188	144161
Kitchener	270	1522	6880	21215	130977	241069
	263	1479	6621	19961	104211	162921
	317	1706	7524	23004	104414	143492
Windsor	326	1730	7765	23862	154825	156032
	289	1624	7258	22460	119550	156432
	339	1872	8213	24790	106961	163603
London	362	1908	8323	25328	113162	142766
	354	1954	8659	26092	208409	188908
	302	1647	7269	21562	114311	193624
St. Catharines	281	1605	7326	22803	295235	323755
	276	1513	6773	20702	123985	220844
	636	3392	14757	44635	168160	291856
Hamilton	689	3547	15526	47796	255346	381554
	662	3472	15172	46608	186617	347964
	881	4570	20097	60957	270496	594061
Ottawa	854	4450	19374	59242	515342	359621
	866	4443	19374	59092	275722	414098

TABLE 3.3. Performance of Phi Statistics

Census Area	Percentage Error					
	10%	25%	50%	75%	100%	150%
	1558	3870	7962	12836	20480	28838
Oshawa	1543	3764	7757	12562	19687	26437
	1469	3678	7648	12480	19724	23711
	1344	3374	7025	11498	18531	24805
Thunder Bay	1440	3538	7403	12141	19485	23530
	1590	3985	3282	13525	21550	29490
	3717	9165	18988	30933	48246	59168
Kitchener	3613	9000	18860	31148	50603	68778
	3602	8872	18344	29619	46505	59782
	3763	9255	19050	30870	47741	57126
Windsor	3793	9240	19174	31190	49046	57358
	3562	8929	18572	30479	47489	57975
	4427	10852	22319	36260	57453	67426
London	4332	10544	21616	35094	53566	64347
	4352	10789	22179	35840	56244	67550
	3891	9558	19734	31966	50082	64819
St. Catharines	3715	9302	19649	32810	57638	67019
	3690	9065	18771	30622	48748	64804
	7409	18322	37578	60780	89723	114609
Hamilton	7759	18839	38662	62922	97032	114235
	7591	18592	38153	61921	91686	115900
	10024	24509	50463	81816	123341	156235
Ottawa	9845	24039	49247	79757	120985	151373
	9974	24193	49549	80154	122776	150706

TABLE 3.4. Performance of Likelihood Statistics

Census Area	Percentage Error 10%	25%	50%	75%	100%	150%
	51	309	1308	3411	6689	11379
Oshawa	54	307	1292	3258	6305	10349
	49	292	1243	3190	6092	10224
	46	265	1135	2995	7461	11608
Thunder Bay	52	291	1250	3255	8015	10452
	57	334	1416	3669	8924	14842
	122	747	3171	8250	19073	25389
Kitchener	116	728	3165	8453	21090	34025
	113	700	2984	7697	18434	26670
	136	810	3372	8692	19637	24574
Windsor	142	817	3443	8887	20609	25257
	126	772	3277	8610	19496	26060
	145	887	3721	9350	22832	29021
London	155	896	3715	9103	20915	26705
	153	918	3841	9827	23096	29604
	146	795	3306	8488	19940	29703
St. Catharines	136	781	3392	9169	26642	32733
	134	731	3073	8051	19678	29967
	354	1675	6622	16812	33898	49314
Hamilton	397	1746	6939	17798	39579	49066
	379	1701	6779	17318	35019	50302
	175	2012	8733	22648	47873	69530
Ottawa	162	1851	8420	21799	47358	65875
	192	1974	8382	21863	48636	65064

TABLE 3.5. Performance of Standard Deviation of Residuals and Mean Absolute Error

Cenasus Area	Percentage Error 10%		25%		50%		75%		100%		150%	
	6.8	3.2	17.1	7.9	34.2	15.8	51.7	23.9	70.6	32.6	97.9	43.8
Oshawa	6.4	3.2	15.9	7.7	31.9	15.5	48.6	23.7	67.3	32.8	82.6	39.8
	6.0	3.0	15.2	7.5	30.8	15.1	47.7	23.2	67.3	32.1	80.2	38.2
	4.2	2.1	10.6	5.3	21.6	10.7	33.3	16.6	46.9	23.2	60.0	30.0
Thunder Bay	4.4	2.3	10.7	5.5	21.9	11.2	33.7	17.3	47.3	24.1	58.7	29.5
	5.0	2.5	12.6	6.3	25.6	12.6	39.4	19.3	54.6	26.6	67.4	33.3
	4.3	1.8	10.7	4.4	21.8	8.9	33.5	13.6	46.5	18.6	58.0	22.9
Kitchener	4.1	1.8	10.5	4.4	22.1	8.9	35.9	14.0	55.6	19.8	62.5	24.8
	4.1	1.8	10.4	4.3	20.9	8.6	31.9	13.1	44.2	17.8	56.2	22.6
	2.5	1.2	6.4	2.9	12.7	5.7	19.3	8.6	26.2	11.7	31.5	14.4
Windsor	2.6	1.2	6.5	2.9	13.2	5.7	20.4	8.7	28.6	11.9	31.6	14.1
	2.4	1.1	6.2	2.8	12.6	5.6	19.6	8.6	27.2	11.8	32.1	14.4
	3.6	1.3	9.2	3.1	19.2	6.1	31.4	9.4	56.7	13.5	43.7	15.2
London	3.1	1.2	7.8	2.9	15.8	5.9	24.4	8.9	34.2	12.2	40.8	15.1
	3.1	1.2	7.9	3.0	16.0	6.0	24.5	9.1	34.1	12.3	38.3	14.8
	3.9	1.4	9.8	3.3	19.8	6.7	30.5	10.3	42.7	14.2	55.5	11.6
St. Catharines	3.8	1.3	10.0	3.3	21.5	6.7	35.5	10.6	54.6	15.4	59.7	17.8
	3.6	1.3	9.1	3.2	18.8	6.4	29.5	9.8	42.2	13.5	76.7	18.5
	1.9	0.6	4.8	1.5	9.7	3.0	14.7	4.5	20.1	6.1	25.4	7.6
Hamilton	2.1	0.6	5.2	1.5	10.6	3.1	16.4	4.6	22.9	6.3	25.7	7.5
	2.0	0.6	5.0	1.5	10.0	3.0	15.3	4.5	21.1	6.1	23.6	7.6
	2.4	0.7	5.9	1.7	11.9	3.3	18.4	5.0	26.1	6.8	30.7	8.3
Ottawa	2.3	0.7	5.7	1.6	11.5	3.2	17.5	4.9	24.1	6.6	29.3	8.2
	2.4	0.7	6.0	1.6	12.2	3.2	19.0	4.9	27.6	6.7	29.2	8.2

Cell entries are Standard Deviation of Residuals and Mean Absolute Error

CHAPTER 4

AGGREGATE TRIP DISTRIBUTION MODELS

This chapter examines the goodness of fit changes for a range of aggregate gravity model types. The explanatory qualities of production constrained, attraction constrained and doubly constrained versions of the gravity model using distance as the inter-zonal cost measure are first examined. Changes in the calibration qualities are then examined when travel times and a combination of travel time and travel distance are used respectively as the cost measures. A variety of stratified trip distribution models are examined in Chapter 5.

4.1 Production Constrained Models

The previous studies of the census journey to work data [1] have shown that the doubly constrained version of the gravity model was best in simulating the observed home to work linkages. Production constrained and attraction constrained versions of the gravity model have been estimated for eight Ontario census areas in order to provide a benchmark against which some of the modifications to the gravity model described in Chapter 5 may be judged.

The basic form of the production constrained gravity model estimated in this study is:

$$T_{ij}^{*} = A_i \, O_i \, D_j \exp(-\beta C_{ij}) \qquad (24)$$

where $$A_i = [\sum_j D_j \exp(-\beta C_{ij})]^{-1} \qquad (25)$$

and the terms have been defined previously in Chapter 3.

Table 10 summarizes the characteristics of the production constrained models estimated for eight Ontario census areas. This table shows that the observed and simulated mean trip lengths are very close to each other in spite of the fact that the calibration criterion was minimization of the sum of the absolute differences between the ordinates of the observed and simulated trip length frequency distributions. The β magnitudes range from 0.152 in Windsor to 0.237 in the Kitchener census area and tend to decrease in magnitude with increasing mean trip length. The transformed phi magnitudes shown in the right hand column of Table 10 suggest that the goodness of fit of the production constrained models is equivalent to a randomly introduced error of 75-100 percent.

4.2 Attraction Constrained Models

The basic form of the attraction constrained model estimated in this study is:

$$T_{ij}^{*} = B_j O_i D_j \exp(-\beta C_{ij}) \qquad (26)$$

where $$B_j = [\sum_i O_i \exp(-\beta C_{ij})]^{-1}$$

The properties of the attraction constrained gravity models for eight census areas are summarized in Table 11. The β magnitudes are all smaller than the β magnitudes of the production constrained model. The β magnitudes tend to decrease with increasing mean trip length but at a faster rate than for the production constrained models. The phi statistic and the transformed phi indicate that the goodness of fit of the attraction constrained models is not as high as for the production constrained models.

TABLE 10. Properties of Production Constrained Gravity Models for Eight Census Areas

Census Area	β	Mean Trip Length (kms) Obs.	Mean Trip Length (kms) Sim.	R^2	Phi	$\frac{Phi}{T}$
Oshawa	0.177	4.10	4.24	0.84	12,963	0.38
Thunder Bay	0.191	5.37	5.19	0.87	11,142	0.34
Kitchener	0.237	4.80	4.70	0.85	33,746	0.42
Windsor	0.152	7.59	7.39	0.75	36,900	0.49
St.Catharines	0.210	6.50	6.11	0.94	41,801	0.46
London	0.177	6.57	6.12	0.94	54,308	0.56
Hamilton	0.160	8.05	8.23	0.76	87,294	0.57
Ottawa	0.202	7.63	7.48	0.74	130,284	0.65

TABLE 11. Properties of Attraction Constrained Gravity Models for Eight Census Areas

Census Area	β	Mean Trip Length (kms) Obs.	Mean Trip Length (kms) Sim.	R^2	Phi	$\frac{Phi}{T}$
Oshawa	0.138	4.10	4.22	0.83	13,720	0.40
Thunder Bay	0.172	5.37	5.13	0.84	12,563	0.38
Kitchener	0.230	4.80	4.69	0.86	34,439	0.42
Windsor	0.084	7.59	7.75	0.62	43,829	0.59
St.Catharines	0.201	6.50	6.06	0.94	44,385	0.48
London	0.146	6.57	6.60	0.92	53,358	0.55
Hamilton	0.102	8.05	.57	0.58	103,136	0.68
Ottawa	0.122	7.63	7.57	0.68	151,152	0.75

4.3 Doubly Constrained Models

The doubly constrained gravity model estimated in this study has been defined in equation (3). The model has been estimated for the fifteen Ontario census areas and the properties of the calibrated models are summarized in Table 12. The observed and simulated mean trip lengths are close to each other for most census areas. The phi and transformed phi statistics shown in Table 12 indicate that the doubly constrained gravity model is superior to both the production constrained and attraction constrained versions of the model. While the β parameter tends to decrease with increasing mean trip length there is a great deal of variability in the relationship. The β parameter magnitudes for the doubly constrained model tend to be larger than the parameter magnitudes for the production constrained models.

4.4 Travel Time and Time-Distance Trip Cost Functions

Doubly constrained models have been calibrated for the Kitchener, Hamilton and Ottawa census areas using network travel times obtained from transport studies performed in these communities. The properties of these time based models are summarized in Table 13. This table indicates that the observed and simulated mean trip lengths are close to each other with the largest difference being for the Kitchener CMA. The phi statistics indicate that the calibrated models are marginally inferior to those calibrated using network distances. It should be noted however that the difference decreases as the city size increases. This would indicate that distance-based networks are more effective for smaller cities and time-based networks more effective for larger cities with the cross-over point being in the range of 600,000 to 1,000,000 people.

TABLE 12. Properties of Doubly Constrained Gravity Models for Fifteen Census Areas Using Distance Cost Functions

Census Area	β	Mean Trip Length (kms) Obs.	Sim.	R^2	Phi	$\frac{Phi}{T}$
Guelph	0.165	3.76	3.80	0.86	5,742	0.29
Peterborough	0.288	3.30	3.11	0.92	5,709	0.28
Brantford	0.206	4.20	4.13	0.82	7,114	0.29
Sarnia	0.091	6.33	6.42	0.92	6,485	0.27
Kingston	0.105	5.68	5.86	0.90	8,281	0.31
Sault Ste.Marie	0.131	4.59	4.77	0.94	6,566	0.27
Oshawa	0.213	4.10	4.21	0.87	11,008	0.33
Thunder Bay	0.201	5.37	5.09	0.87	10,775	0.33
Sudbury	0.120	8.63	8.73	0.75	20,876	0.46
Kitchener	0.259	4.80	4.60	0.90	29,924	0.37
St.Catharines	0.214	6.50	6.05	0.96	37,359	0.41
Windsor	0.115	7.59	7.91	0.77	34,625	0.47
London	0.180	6.57	6.40	0.95	44,360	0.46
Hamilton	0.159	8.05	8.26	0.81	83,017	0.55
Ottawa	0.212	7.63	7.40	0.79	119,509	0.60

TABLE 13. Properties of Doubly Constrained Gravity Models for Three Census Areas Using Travel Time Cost Functions

Census Area	β	Mean Trip Length (mins) Obs.	Sim.	R^2	Phi	$\frac{Phi}{T}$
Kitchener	0.150	10.23	9.79	0.74	36,212	0.45
Hamilton	0.121	13.38	13.48	0.76	88,430	0.58
Ottawa	0.119	13.51	13.71	0.78	121,352	0.60

The doubly constrained model was also estimated using travel costs which represented various combinations of inter-zonal distance and travel time for the Kitchener census area. The four generalized cost functions used are:

$$\text{I} \qquad c^*_{ij} = \beta_{distance} . c_{ij} + \beta_{time} . tt_{ij} \tag{27}$$

$$\text{II} \qquad c^*_{ij} = \beta_{distance} . c_{ij} . \text{FACT} + \beta_{time} . tt_{ij} . (1 - \text{FACT}) \tag{28}$$

$$\text{III} \qquad c^*_{ij} = c_{ij}/\text{Mean Trip Distance} + tt_{ij}/\text{Mean Trip Time} \tag{29}$$

$$\text{IV} \qquad c^*_{ij} = [c_{ij}/\text{Mean Trip Distance}].\text{FACT} + [tt_{ij}/\text{Mean Trip Time}]\ (1 - \text{FACT}) \tag{30}$$

where c^*_{ij} = "generalized" trip cost between zones i and j

$\beta_{distance}$ = magnitude estimated for the doubly constrained model using network distances

β_{time} = magnitude estimated for the doubly constrained model using network travel time

c_{ij} = network travel distance between zones i and j

tt_{ij} = network travel times between zones i and j

FACT is an index used to weight distance and speed differentially depending on the average travel speed between zones i and j

FACT = 1 if average speed $\geq$ 80 km/hr

$= \dfrac{\text{average speed} - 10}{70}$ if $10 \leq \text{speed} \leq 80$ km/hr

= 0 if average speed $\leq$ 10 km/hr

With the generalized cost function specified in equation (27) the deterrence effects of distance and time are equally weighted. In equation (28) when the average speed is greater than 80 km/hr the deterrence effect is assumed to be all distance and when the average speed is less than 10 km/hr the deterrence effect is assumed to be all travel time. In equation (29) the distance and speed components are effectively normalized because they are divided by the area-wide mean trip distances and mean trip length and equation (30) has the same effect as equation (28).

Table 14 summarizes the properties of the doubly constrained gravity models using the four different cost functions. These parameter magnitudes cannot be compared directly since the generalized travel cost units vary between each model. The observed and simulated mean trip lengths are close to each other but there is little change in the overall goodness of fit of the doubly constrained model. The models are of better quality than those using travel times but are marginally inferior to that estimated for the Kitchener census area using network travel distances as the travel costs.

4.5 Comparisons of Residuals Across Model Types

It is also useful to examine the trip interchange residuals in some detail in addition to the goodness of fit statistics discussed in the previous sections. The trip interchange residuals for the Ottawa census area are examined for four separate model types in this section. The spatial distributions of residuals observed for Ottawa are broadly representative of the residuals for all census areas.

Figures 60 and 61 show the over-estimation (simulated > observed) and under-estimation (simulated < observed) residuals for the

TABLE 14. Properties of the Doubly Constrained Gravity Model for the Kitchener Census Area Using Different Cost Functions

Cost Function	β	Mean Trip Length		R^2	Phi	$\frac{Phi}{T}$
		Obs.	Sim.			
I	0.490	2.77	2.72	0.85	31,123	0.38
II	0.998	1.50	1.46	0.82	32,366	0.40
III	0.300	4.80	4.56	0.85	31,563	0.39
IV	0.303	4.98	4.76	0.84	31,833	0.39

FIGURE 60. Overestimation Residuals for Production Constrained Gravity Model in Ottawa

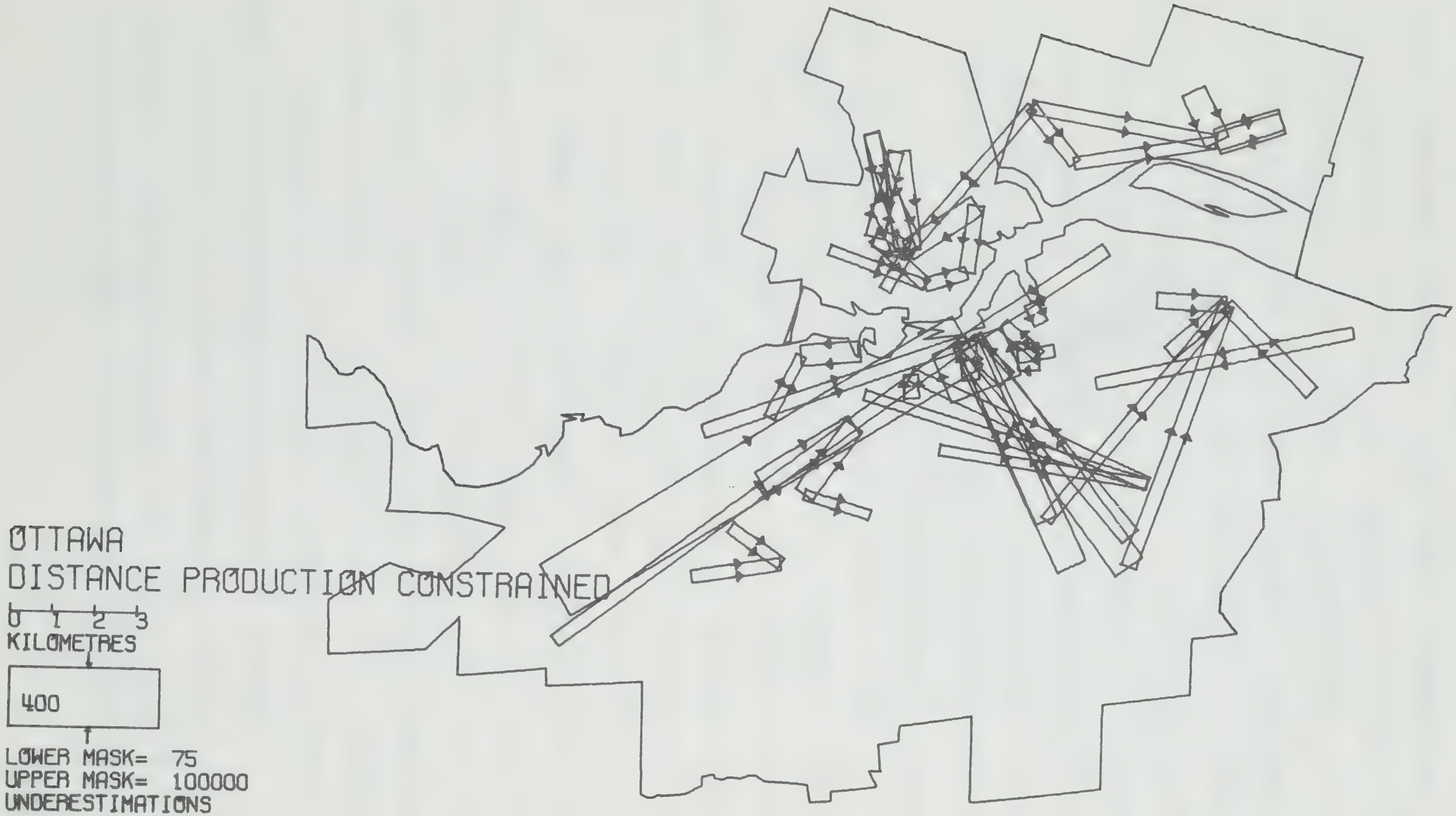


FIGURE 61. Underestimation Residuals for Production Constrained Gravity Model in Ottawa

production constrained gravity model estimated for Ottawa. Trips from the inner suburbs of Ottawa and the residential areas in Hull to the Ottawa central area employment zones tend to be over-estimated by the production constrained model. This tendency to over-estimate trips to the central area is directly related to the heavy concentrations of employment in the Ottawa CBD. The under-estimation residuals are primarily associated with trips to suburban employment locations in both Ottawa and Hull although there are some under-estimation residuals associated with the longer trips to the CBD. This latter type of residual is probably associated with the timing of development of certain residential sub-divisions and the growth of Federal Government job opportunities in the central area.

Figures 62 and 63 illustrate the over-estimation and under-estimation residuals for the attraction constrained gravity model estimated for Ottawa. The over-estimation residuals tend to be clustered around the central employment area reflecting the tendency of the gravity model to allocate trips to the closest available opportunities. The under-estimation residuals are almost all associated with the long trips to the CBD employment opportunities and these residuals are a major cause of the poorer goodness of fit of the attraction constrained model. The frequency distribution of the labour force by census tract is much more uniform than employment and inspection of the β-parameter magnitudes in Tables 10 and 11 shows that the β-parameter magnitude for the attraction constrained model falls in order to compensate.

Figures 64 and 65 show the over-estimation and under-estimation residuals for the doubly constrained version of the gravity model. Figure 64 shows that the doubly constrained model tends to reduce the number of over-estimation residuals to the central area compared with

FIGURE 62. Overestimation Residuals for Attraction Constrained Gravity Model in Ottawa

FIGURE 63. Underestimation Residuals for Attraction Constrained Gravity Model in Ottawa

FIGURE 64. Overestimation Residuals for Doubly Constrained Gravity Model in Ottawa

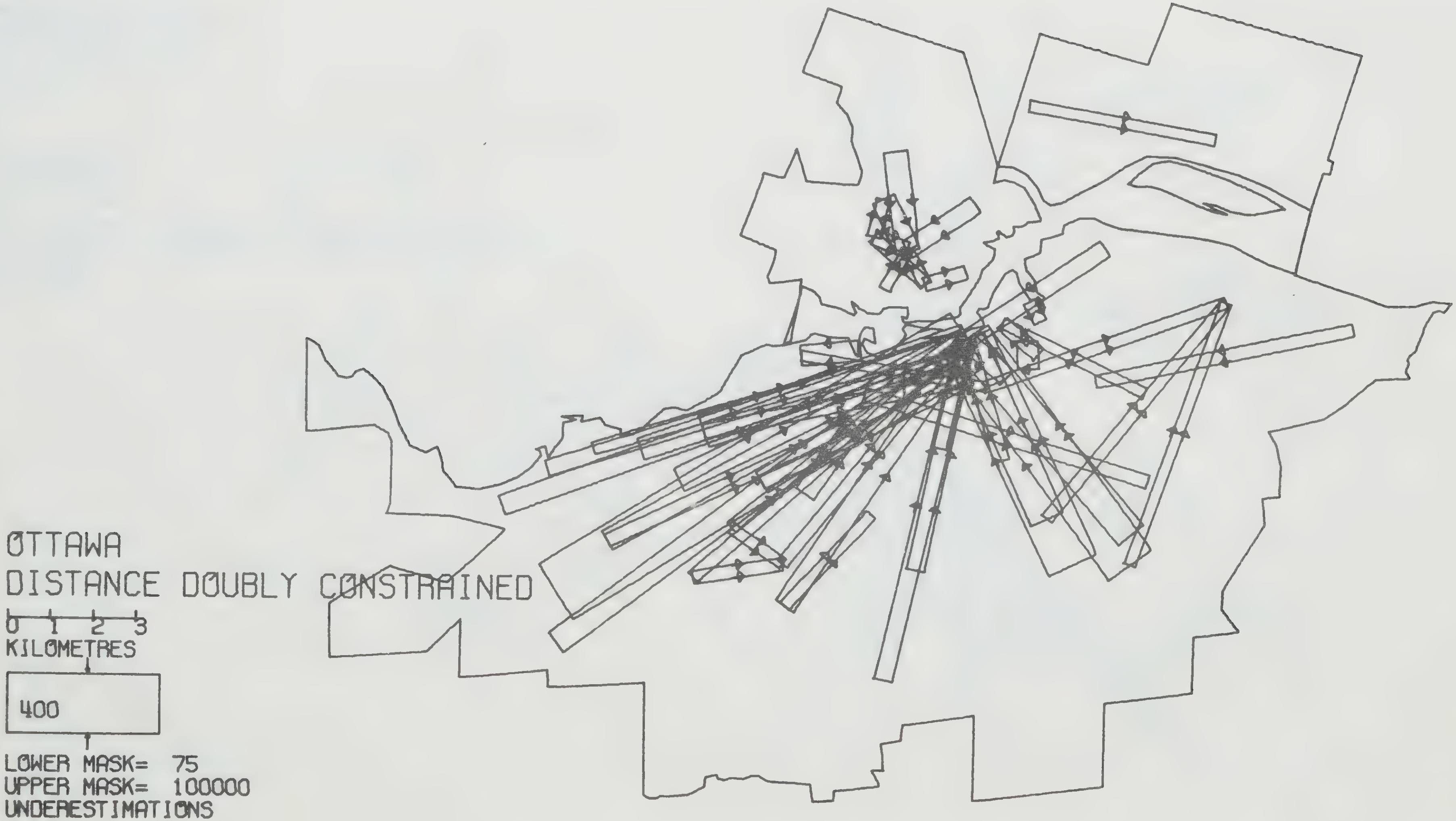


FIGURE 65. Underestimation Residuals for Doubly Constrained Gravity Model in Ottawa

the production constrained model. On the other hand Figure 65 shows that the under-estimation residuals tend to increase for the doubly constrained model compared with the production constrained model.

Figures 66 and 67 show the over-estimation and under-estimation residuals for the doubly constrained version of the gravity model using network travel times rather than network travel distances.

A comparison of the over-estimation residuals plotted in Figure 66 with those plotted in Figure 64 illustrates the tendency of the model to further over-estimate trips from some of the outer residential areas of Hull. Comparisons of the under-estimation residuals in Figures 67 and 65 show that there is little change in the spatial pattern of residuals.

The relative importance of the over-estimated and under-estimated residuals is further illustrated in Table 15 where the phi-statistics are summarized for the four gravity model types. This table demonstrates quite clearly that the major source of error is the under-estimation of trip interchange magnitudes and that the superior behaviour of the doubly constrained model is due to a reduction in the under-estimation residuals. The table also demonstrates that the intra-zonal under-estimation residuals are a significant proportion of the total. However, in reviewing this decomposition of the residuals it should be recalled that the phi-statistic is more sensitive to under-estimation errors than to over-estimation errors.

4.6 Comparison of Residuals for Alternative Cost Functions

A more detailed analysis of the doubly constrained gravity models calibrated for the Kitchener census area using alternative travel

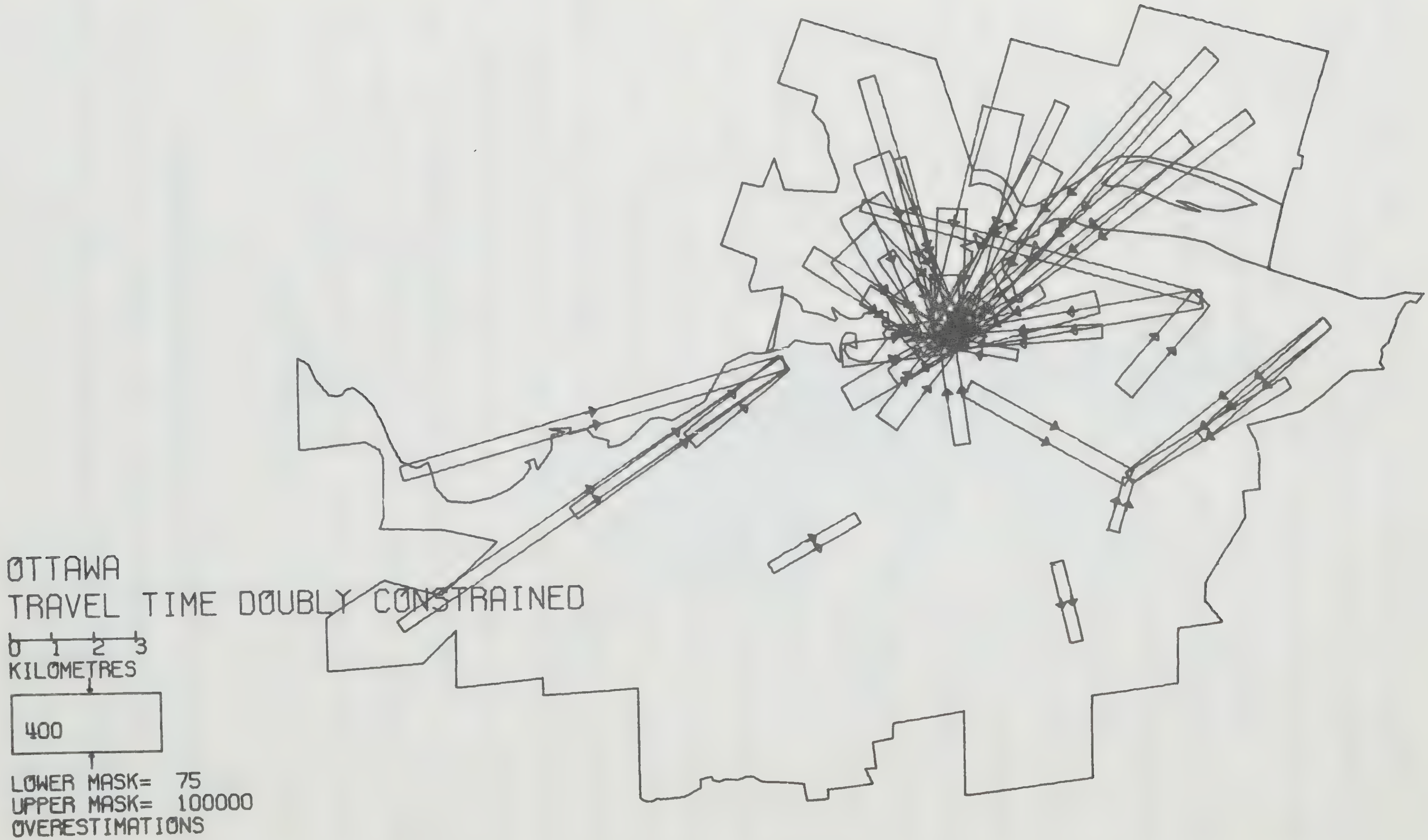


FIGURE 66. Overestimation Residuals for Doubly Constrained Gravity Model in Ottawa Using Travel Times

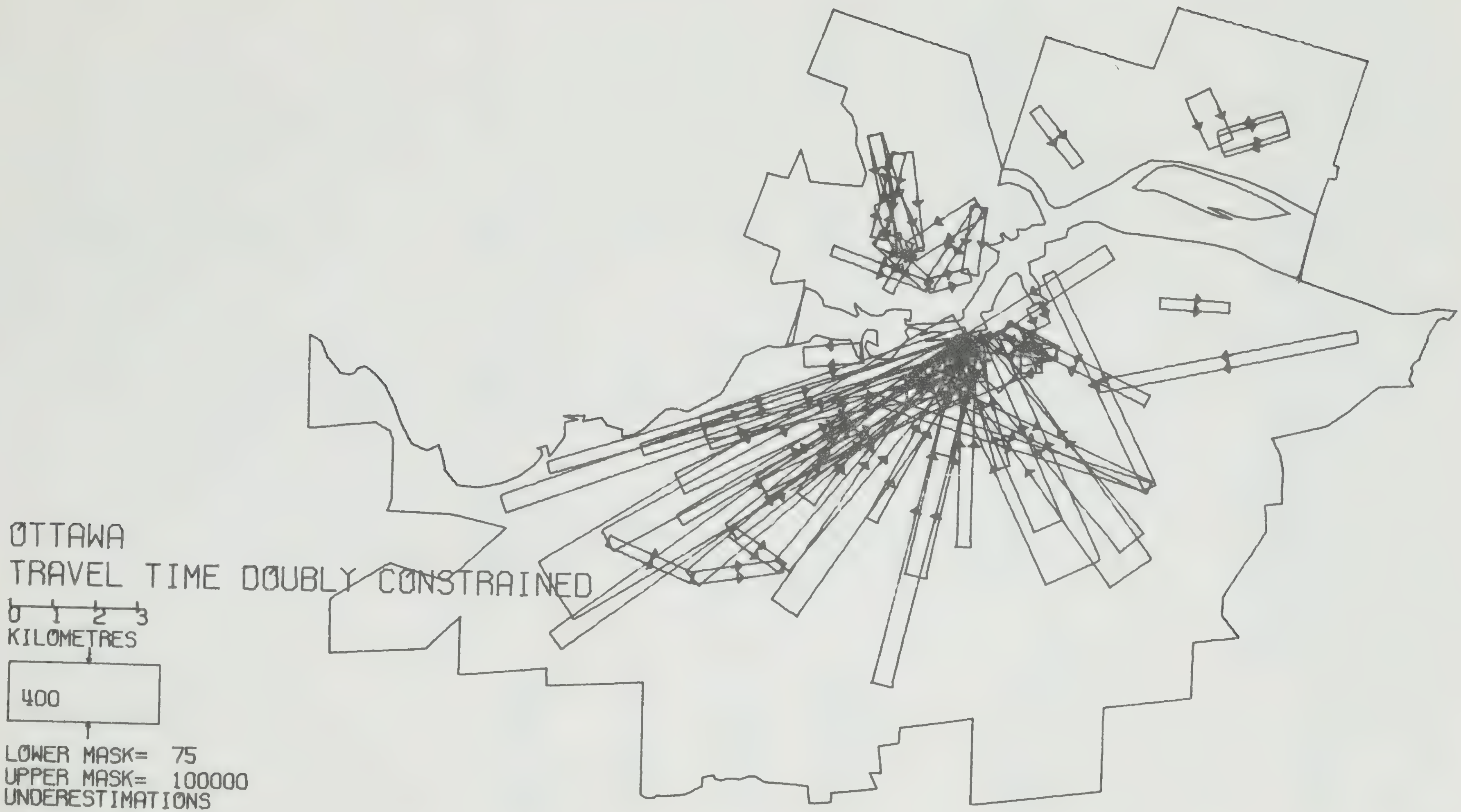


FIGURE 67. Underestimation Residuals for Doubly Constrained Gravity Model in Ottawa Using Travel Times

TABLE 15. Relative Importance of Over- and Under-Estimation Residuals for Four Model Types for Ottawa

Gravity Model Type	Total Phi		Intra-Zonal Phi	
	Over-Est. Residuals	Under-Est. Residuals	Over-Est. Residuals	Under-Est. Residuals
Production Constrained	23,634	106,650	426	21,597
Attraction Constrained	21,717	129,435	0	30,055
Doubly Constrained	21,049	98,460	873	16,622
Doubly Constrained With Time	20,814	100,537	955	19,189

cost functions is shown in Table 16. This table shows that there is little change between the different travel cost function types with the distance and distance-time travel cost functions providing the best goodness of fit.

TABLE 16. Goodness of Fit Statistics for Doubly Constrained Gravity Models Using Different Travel Cost Functions for Kitchener Census Area

Travel Cost Function Type	Total Phi		Intra-Zonal Phi	
	Over-Est. Residuals	Under-Est. Residuals	Over-Est. Residuals	Under-Est. Residuals
Distance	8,178	22,875	485	3,008
Time	8,701	23,720	1,260	3,868
I	8,247	23,007	909	3,417
II	8,646	23,107	1,131	3,553
III	8,556	21,745	1,192	2,943
IV	8,752	27,510	1,173	3,231

CHAPTER 5

STRATIFIED TRIP DISTRIBUTION MODELS

This chapter examines the qualities of a number of stratified gravity models. The first group of stratified models examined in this chapter are multi-parameter gravity models in which separate β-parameter magnitudes are estimated for each origin zone in production constrained, attraction constrained and doubly constrained forms of the gravity model. The second group of stratified gravity models discussed have separate β-parameter magnitudes estimated for from four to eight calibration sub-regions where the census tracts in each of these calibration sub-regions have been identified by the clustering procedure described in Chapter 3. The final set of gravity models evaluated in this chapter are stratified by socio-economic group.

5.1 Multi-Parameter Production Constrained Models

Table 17 summarizes the characteristics of the multi-parameter production constrained gravity models estimated for eight census areas. With these models separate β-parameter magnitudes were estimated for each residential zone. The table shows that in all cases the area-wide observed and simulated mean trip lengths are close to each other for all census areas with the multi-community census areas of Kitchener and St. Catharines exhibiting the largest differences between the observed and simulated mean trip lengths. A comparison of the goodness of fit statistics in Table 17 with those presented in Table 10 for the single

TABLE 17. Properties of the Multi-Parameter Production Constrained Gravity Models for Eight Census Areas

Census Area	β Range	Mean Trip Length (kms) Obs	Mean Trip Length (kms) Sim	R^2	Phi	$\frac{Phi}{T}$
Oshawa	.003 - .472	4.10	4.16	0.90	10,827	0.32
Thunder Bay	.032 - .314	5.37	5.21	0.88	10,775	0.33
Kitchener	.077 - .479	4.80	4.59	0.89	30,420	0.37
Windsor	.015 - .414	7.59	7.52	0.83	33,903	0.46
St. Catharines	.127 - .367	6.50	5.97	0.97	36,164	0.39
London	.030 - .497	6.57	6.42	0.95	49,043	0.50
Hamilton	.032 - .489	8.05	7.92	0.87	80,669	0.53
Ottawa	.047 - .497	7.63	7.33	0.75	126,048	0.63

parameter production constrained models shows that the multi-parameter models all exhibited superior goodness of fit characteristics. The largest improvements were in the Kitchener and St. Catharines census areas.

Figure 68 shows the spatial distribution of origin-zone specific β-parameters for the Kitchener census area. It should be recalled that a β-parameter magnitude of 0.237 was obtained for the aggregate production constrained model while the range obtained for the multi-parameter model was from .077 to 0.479. Figure 68 illustrates that most of the fringe residential zones not located close to employment areas have lower than average β magnitudes while those for the inner residential areas closer to the large employment zones have much larger values. The calibrated β magnitudes for each residential zone reflects not only the mean trip length characteristics of the zone but the proximity of the zone to large concentrations of trip attractions.

Figure 69 shows the spatial distribution of origin-zone specific β-parameter magnitudes for the London census area. The β magnitude for the aggregate production constrained model was 0.177 and the range illustrated in Figure 69 is from 0.030 to 0.497. Once again the lowest magnitudes are on the urban fringe and the largest values in zones close to the CBD and the University of Western Ontario.

5.2 Multi-Parameter Attraction Constrained Models

Table 18 summarizes the properties of the multi-parameter attraction constrained models estimated for eight Ontario census areas. A comparison of the goodness of fit statistics in Table 18 with those

FIGURE 68. β-Parameter Distribution in Kitchener

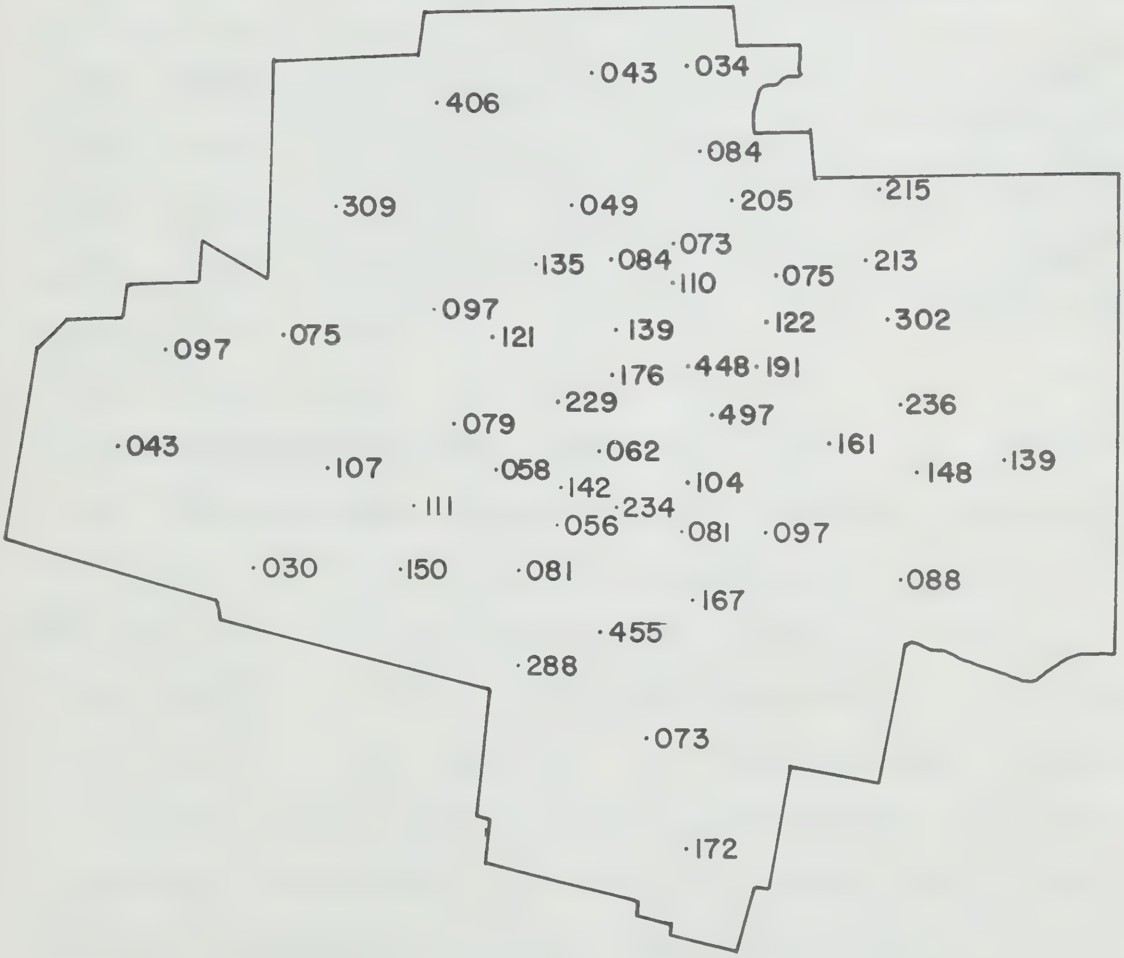


FIGURE 69. β-Parameter Distribution in London

TABLE 18. Properties of Multi-Parameter Attraction Constrained Gravity Models for Eight Census Areas

Census Area	β Range	Mean Trip Length (kms) Obs	Sim	R^2	Phi	$\frac{Phi}{T}$
Oshawa	.073 - .414	4.10	4.04	0.73	18,004	0.53
Thunder Bay	.118 - .497	5.37	5.21	0.84	11,717	0.35
Kitchener	.127 - .492	4.80	4.61	0.85	33,562	0.41
Windsor	.045 - .497	7.59	7.29	0.64	44,577	0.60
St. Catharines	.103 - .409	6.50	5.69	0.96	41,108	0.45
London	.073 - .497	6.57	6.49	0.89	56,708	0.58
Hamilton	.073 - .497	8.05	7.58	0.58	112,857	0.74
Ottawa	.073 - .497	7.63	6.87	0.68	144,116	0.72

in Table 11 indicates that the multi-parameter attraction constrained model has poorer goodness of fit characteristics than the aggregate model. The only census areas in which improvements occur are Thunder Bay, Kitchener and St. Catharines, the census areas with strong multi-community spatial structures.

5.3 Multi-Parameter Doubly Constrained Models

Table 19 summarizes the properties of the multi-parameter doubly constrained models estimated for the eight Ontario census areas. A comparison of the goodness of fit statistics of Table 19 with those of Table 12 indicates that the multi-parameter doubly constrained model is superior to the single parameter doubly constrained model. In addition the doubly constrained multi-parameter model is superior to the other two multi-parameter models.

5.4 Sub-Region Specific Models

Table 20 summarizes the properties of production constrained gravity models estimated with separate β-parameters for a number of individual calibration sub-regions for eight census areas. The cluster analyses described in Chapter 2 formed the basis for the calibration sub-regions isolated for each census area. This table shows that the observed and simulated mean trip lengths are close together for all census areas. A comparison of the goodness of fit statistics in Table 20 with those in Tables 10, 12 and 17 show that the sub-region specific model is superior to the single parameter production constrained model and inferior to both the single parameter doubly constrained and multi-parameter production constrained models.

TABLE 19. Properties of Multi-Parameter Doubly Constrained Gravity Models for Eight Census Areas

Census Area	β Range	Mean Trip Length (kms) Obs	Mean Trip Length (kms) Sim	R^2	Phi	$\frac{Phi}{T}$
Oshawa	.056 - .461	4.10	4.16	0.91	10,204	0.30
Thunder Bay	.094 - .354	5.37	5.28	0.88	9,922	0.30
Kitchener	.118 - .438	4.80	4.54	0.92	27,918	0.34
Windsor	.034 - .480	7.59	7.63	0.84	32,521	0.44
St. Catharines	.118 - .420	6.50	6.09	0.97	34,102	0.37
London	.072 - .491	6.57	6.36	0.96	43,133	0.44
Hamilton	.045 - .483	8.05	7.97	0.88	79,285	0.52
Ottawa	.099 - .493	7.63	7.38	0.83	113,143	0.56

TABLE 20. Properties of Sub-Region Specific Parameter Production Constrained Gravity Models of Eight Census Areas

Census Area	β No.	β Range	Mean Trip Length (kms) Obs	Mean Trip Length (kms) Sim	R^2	Phi	$\frac{Phi}{T}$
Oshawa	4	.046 - .369	4.10	4.29	0.88	11,434	0.34
Thunder Bay	3	.129 - .206	5.37	5.20	0.87	11,293	0.34
Kitchener	3	.182 - .313	4.80	4.55	0.87	33,023	0.41
Windsor	5	.098 - .414	7.59	7.44	0.75	36,589	0.49
St. Catharines	3	.188 - .223	6.50	5.95	0.95	41,223	0.45
London	4	.073 - .204	6.57	6.61	0.94	51,310	0.53
Hamilton	7	.058 - .229	8.05	8.03	0.82	86,310	0.57
Ottawa	8	.139 - .407	7.63	7.44	0.74	125,013	0.62

Table 21 summarizes the results obtained from the calibration of doubly constrained models incorporating specific β-parameters for the same calibration sub-regions used for the production constrained models summarized in Table 20. A comparison of the goodness of fit statistics of Table 21 with those in Tables 12 and 19 show that the doubly constrained sub-region specific models performed at about the same level or marginally better than the single parameter doubly constrained models but were inferior to the multi-parameter doubly constrained models.

5.5 Models Stratified by Socio-Economic Group

A number of gravity models have been estimated using the stratified journey to work tabulations obtained from Statistics Canada and described in reference [1]. The three two-way classifications obtained are:

Automobile Ownership

The "number of automobiles in a household" variable formed the basis for this two-way classification where the following groups were identified:

Group 1 - no automobile
- members of labour force from households with 0 automobiles
- members of labour force who are non-heads of household from a household with 1 automobile

Group 2 - automobile available
- heads of households from a 1 automobile household
- members of labour force from households with 2+ automobiles

Tenancy

Group 1 - owner occuped dwelling unit
Group 2 - rental dwelling unit

Period of Residence

Group 1 - individuals residing in current residence for more than 5 years
Group 2 - individuals residing in current residence for five years or less

TABLE 21. Properties of Sub-Region Specific Parameter Doubly Constrained Gravity Models of Eight Census Areas

Census Area	β No.	β Range	Mean Trip Length (kms) Obs	Mean Trip Length (kms) Sim	R^2	Phi	$\frac{Phi}{T}$
Oshawa	4	.152 - .377	4.10	4.03	0.88	11,051	0.33
Thunder Bay	3	.073 - .223	5.37	5.44	0.87	10,127	0.31
Kitchener	3	.249 - .292	4.80	4.50	0.90	30,368	0.37
Windsor	5	.115 - .497	7.59	7.70	0.77	34,594	0.46
St. Catharines	3	.191 - .309	6.50	5.85	0.96	38,384	0.42
London	4	.155 - .208	6.57	6.48	0.94	44,273	0.46
Hamilton	7	.118 - .191	8.05	8.16	0.83	83,120	0.54
Ottawa	8	.116 - .309	7.63	7.60	0.80	116,139	0.58

Table 22 summarizes the characteristics of production constrained models estimated for members of the labour force without a car available and with a car available for the journey to work. The only goodness of fit statistics reported in Table 22 are the coefficients of determination since these production constrained models were calibrated prior to the beginning of the research project described in this report. Inspection of Table 22 shows that there is generally good agreement between the observed and simulated mean trip lengths for the two groups and that in most census areas the β-parameter magnitude for the non-car owners is larger than for car owners.

Table 23 summarizes the characteristics of doubly constrained models estimated for non-car owners in three census areas using both network distances and network times. Table 24 summarizes the information for the models calibrated for car owners. Since the zone systems and numbers of linkages are fewer for the stratified models it is appropriate to compare the transformed phi statistics in Tables 23 and 24 with those in Table 12. This comparison would suggest that the doubly constrained models stratified by car ownership status perform at a superior level to the doubly constrained single parameter models. The β-parameter magnitudes for the captive members of the labour force are significantly higher than for the non-captives for all three census areas reflecting the greater sensitivity of captives to travel costs.

Tables 25, 26 and 27 summarize the characteristics of the stratified models calibrated for home owners and home renters. Tables 28, 29 and 30 summarize the characteristics of the stratified models calibrated for periods of residence of greater than six years and for five years and less. The comparison of home owner and home renter behaviour indicates

TABLE 22. Production Constrained Gravity Models for No Car Available and Car Available Groups for Fifteen Census Areas

Census Area	Group	β	Mean Trip Length (km) Obs	Sim	R^2
Guelph	No Car	0.10	3.40	3.60	0.73
	Car	0.14	3.90	3.90	0.78
Peterborough	No Car	0.33	3.00	2.80	0.80
	Car	0.22	3.40	2.24	0.88
Sarnia	No Car	0.08	4.50	5.10	0.68
	Car	0.05	6.70	6.98	0.87
Brantford	No Car	0.32	3.50	3.24	0.80
	Car	0.25	4.40	4.08	0.77
Sault Ste. Marie	No Car	0.23	3.50	3.62	0.79
	Car	0.20	4.90	4.98	0.92
Kingston	No Car	0.07	4.30	4.56	0.83
	Car	0.10	6.00	6.10	0.85
Thunder Bay	No Car	0.26	4.60	4.30	0.89
	Car	0.18	5.60	5.40	0.92
Oshawa	No Car	0.15	3.90	4.10	0.74
	Car	0.22	4.10	4.10	0.80
Sudbury	No Car	0.22	6.40	5.78	0.72
	Car	0.10	9.20	9.20	0.68

TABLE 22. (continued)

Census Area	Group	β	Mean Trip Length (km) Obs	Sim	R^2
Kitchener	No Car	0.32	4.10	3.84	0.83
	Car	0.22	5.10	5.04	0.83
Windsor	No Car	0.17	6.00	6.20	0.76
	Car	0.15	8.00	7.83	0.82
London	No Car	0.22	5.10	5.20	0.90
	Car	0.21	6.90	6.10	0.88
St. Catharines	No Car	0.36	4.70	4.14	0.96
	Car	0.23	6.80	5.90	0.94
Hamilton	No Car	0.25	5.80	5.70	0.82
	Car	0.16	8.80	8.70	0.91
Ottawa	No Car	0.23	6.10	6.30	0.79
	Car	0.17	8.30	8.30	0.75

TABLE 23. Properties of Doubly Constrained Gravity Models Calibrated Using Travel Time and Travel Distance For Three Urban Areas : No Car Available

Census Area and Deterrence Measure	β	Mean Trip Length (kms, mins) Obs	Sim	R^2	Phi	$\frac{Phi}{T}$
Kitchener						
Distance	0.386	4.10	3.63	0.88	10,235	0.36
Time	0.181	9.02	8.87	0.82	11,871	0.42
Hamilton						
Distance	0.237	5.83	5.80	0.87	21,563	0.42
Time	0.169	10.41	10.54	0.81	23,577	0.46
Ottawa						
Distance	0.248	6.15	6.18	0.84	35,789	0.49
Time	0.156	11.66	11.84	0.85	35,355	0.48

TABLE 24. Properties of Doubly Constrained Gravity Models Calibrated Using Travel Time and Travel Distance For Three Urban Areas : Car Available

Census Area and Deterrence Measure	β	Mean Trip Length (kms, mins) Obs	Sim	R^2	Phi	$\frac{Phi}{T}$
Kitchener						
Distance	0.239	5.07	4.91	0.90	16,310	0.29
Time	0.141	10.61	10.48	0.83	19,211	0.34
Hamilton						
Distance	0.147	8.79	9.01	0.97	35,102	0.33
Time	0.113	13.70	14.13	0.84	42,207	0.39
Ottawa						
Distance	0.158	8.34	8.51	0.80	60,604	0.46
Time	0.113	14.31	14.44	0.82	57,245	0.44

TABLE 25. Production Constrained Gravity Models for Home Owner and Home Renter Groups for Fifteen Census Areas

Census Area	Group	β	Mean Trip Length (km) Obs	Mean Trip Length (km) Sim	R^2
Guelph	Owners	0.14	3.60	3.80	0.77
	Renters	0.11	3.80	3.90	0.75
Peterborough	Owners	0.29	3.30	3.10	0.86
	Renters	0.26	3.00	2.80	0.85
Sarnia	Owners	0.07	6.40	6.80	0.82
	Renters	0.02	5.70	5.90	0.85
Brantford	Owners	0.29	4.20	3.90	0.82
	Renters	0.21	3.80	3.60	0.75
Sault Ste. Marie	Owners	0.16	4.60	4.60	0.89
	Renters	0.08	4.10	4.30	0.83
Kingston	Owners	0.13	5.80	5.60	0.92
	Renters	0.08	5.20	5.20	0.81
Thunder Bay	Owners	0.20	5.40	5.20	0.91
	Renters	0.22	4.60	4.40	0.88
Oshawa	Owners	0.23	4.10	4.10	0.78
	Renters	0.16	3.70	3.90	0.80
Sudbury	Owners	0.13	8.60	8.40	0.71
	Renters	0.20	7.90	5.90	0.71

TABLE 25. (continued)

Census Area	Group	β	Mean Trip Length (km) Obs	Sim	R^2
Kitchener	Owners	0.28	4.70	4.40	0.83
	Renters	0.26	4.70	4.30	0.84
Windsor	Owners	0.15	7.70	7.60	0.81
	Renters	0.16	6.70	6.30	0.80
London	Owners	0.21	6.60	6.20	0.88
	Renters	0.19	5.90	5.30	0.91
St. Catharines	Owners	0.28	6.30	5.30	0.95
	Renters	0.24	6.10	5.30	0.93
Hamilton	Owners	0.19	8.30	8.00	0.90
	Renters	0.19	6.90	6.70	0.84
Ottawa	Owners	0.20	8.40	8.20	0.77
	Renters	0.15	6.40	6.80	0.77

TABLE 26. Properties of Doubly Constrained Gravity Models Calibrated Using Travel Time and Travel Distance For Three Urban Areas : Home Owners

Census Area and Deterrence Measure	β	Mean Trip Length (kms, mins) Obs	Sim	R^2	Phi	$\frac{\text{Phi}}{T}$
Kitchener						
Distance	0.310	4.77	4.30	0.89	18,131	0.34
Time	0.162	10.15	9.80	0.81	20,116	0.37
Hamilton						
Distance	0.170	8.26	8.36	0.91	37,954	0.35
Time	0.126	13.13	13.52	0.83	44,016	0.40
Ottawa						
Distance	0.195	8.43	8.29	0.80	55,649	0.49
Time	0.110	14.22	14.95	0.82	50,916	0.45

TABLE 27. Properties of Doubly Constrained Gravity Models Calibrated Using Travel Time and Travel Distance For Three Urban Areas : Home Renters

Census Area and Deterrence Measure	β	Mean Trip Length (kms, mins) Obs	Sim	R^2	Phi	$\frac{Phi}{T}$
Kitchener						
Distance	0.306	4.67	4.14	0.88	10,600	0.35
Time	0.144	9.90	9.97	0.83	11,390	0.38
Hamilton						
Distance	0.191	6.88	6.67	0.86	21,039	0.42
Time	0.123	11.56	12.10	0.80	22,565	0.45
Ottawa						
Distance	0.196	6.43	6.59	0.82	42,722	0.48
Time	0.139	12.24	12.13	0.84	41,002	0.46

TABLE 28. Production Constrained Gravity Models for 6+ Years Resident and 0-5 Years Resident Groups for Fifteen Urban Areas

Census Area	Group	β	Mean Trip Length (km)		R^2
			Obs	Sim	
Guelph	6+ Yrs	0.10	3.50	3.70	0.74
	0-5 Yrs	0.13	3.90	3.90	0.78
Peterborough	6+ Yrs	0.21	3.20	3.20	0.87
	0-5 Yrs	0.18	3.30	3.30	0.84
Sarnia	6+ Yrs	0.12	6.00	5.80	0.85
	0-5 Yrs	0.03	6.40	6.70	0.88
Brantford	6+ Yrs	0.39	4.10	3.40	0.75
	0-5 Yrs	0.22	4.20	3.90	0.79
Sault Ste. Marie	6+ Yrs	0.16	4.50	4.50	0.89
	0-5 Yrs	0.12	4.50	4.60	0.84
Kingston	6+ Yrs	0.19	5.40	5.00	0.89
	0-5 Yrs	0.07	5.60	5.90	0.79
Thunder Bay	6+ Yrs	0.19	5.30	5.20	0.90
	0-5 Yrs	0.21	5.20	4.90	0.91
Oshawa	6+ Yrs	0.23	4.10	4.10	0.79
	0-5 Yrs	0.16	3.90	3.90	0.80
Sudbury	6+ Yrs	0.19	8.20	6.70	0.66
	0-5 Yrs	0.11	8.70	8.50	0.67

TABLE 28. (continued)

Census Area	Group	β	Mean Trip Length (km) Obs	Sim	R^2
Kitchener	6+ Yrs	0.32	4.50	4.00	0.81
	0-5 Yrs	0.24	4.80	4.70	0.87
Windsor	6+ Yrs	0.18	7.10	6.80	0.80
	0-5 Yrs	0.13	7.90	7.90	0.78
London	6+ Yrs	0.23	6.10	5.90	0.87
	0-5 Yrs	0.19	6.60	5.80	0.90
St. Catharines	6+ Yrs	0.29	6.00	5.10	0.95
	0-5 Yrs	0.22	6.10	6.00	0.93
Hamilton	6+ Yrs	0.21	7.90	7.40	0.88
	0-5 Yrs	0.19	7.90	7.60	0.87
Ottawa	6+ Yrs	0.23	7.10	6.90	0.76
	0-5 Yrs	0.16	7.90	8.00	0.78

TABLE 29. Properties of Doubly Constrained Gravity Models Calibrated Using Travel Time and Travel Distance For Three Urban Areas : 6+ Year Residents

Census Area and Deterrence Measure	β	Mean Trip Length (kms, mins) Obs	Mean Trip Length (kms, mins) Sim	R^2	Phi	$\frac{Phi}{T}$
Kitchener						
Distance	0.337	4.57	4.07	0.88	14,059	0.37
Time	0.177	9.87	9.32	0.81	15,150	0.40
Hamilton						
Distance	0.207	7.74	7.44	0.90	31,012	0.39
Time	0.128	12.68	13.24	0.81	34,742	0.43
Ottawa						
Distance	0.223	7.10	7.06	0.81	42,337	0.48
Time	0.135	12.72	13.04	0.22	40,507	0.46

TABLE 30. Properties of Doubly Constrained Gravity Models Calibrated Using Travel Time and Travel Distance For Three Urban Areas : 0-5 Year Residents

Census Area and Deterrence Measure	β	Mean Trip Length (kms, mins) Obs	Sim	R^2	Phi	$\frac{Phi}{T}$
Kitchener						
Distance	0.249	4.83	4.68	0.91	13,187	0.29
Time	0.144	10.18	10.12	0.84	15,837	0.34
Hamilton						
Distance	0.151	7.92	8.13	0.88	29,965	0.38
Time	0.116	12.59	13.08	0.82	32,210	0.41
Ottawa						
Distance	0.164	7.91	8.01	0.81	54,848	0.47
Time	0.119	13.84	13.85	0.82	51,278	0.44

that the differences are not systematic in terms of both the mean trip length comparisons and the β-parameter comparisons. On the other hand the period of residence comparisons show that there are significant differences between the two groups and that the β-parameters in the three census areas for periods of residence of six years and greater are larger than for the shorter periods of residence. The transformed phi statistics show that this stratification has greater explanatory power than the single parameter model and is marginally inferior to the stratification based on car ownership status.

CHAPTER 6

STRATIFIED TRIP GENERATION EQUATIONS

Trip distribution models calibrated by socio-economic group have been described in the previous chapter. In addition reference [1] has described some preliminary trip generation analyses by socio-economic group for the Kitchener CMA. The special journey to work tabulations from Statistics Canada were not obtained for inclusion in reference [1] and this chapter describes the complete set of trip generation regression analyses conducted for all of the Ontario census areas.

The first sections of the chapter summarize the trip generation and trip length characteristics of the socio-economic groups by residential tenancy and period of residence. The subsequent sections provide a more detailed analysis of the travel characteristics of non-car owners and car owners.

6.1 Travel Characteristics by Tenancy and Period of Residence

Table 31 summarizes the work trip linkage rates for the fifteen Ontario urban areas for dwelling unit owners and renters. The entries in this table are based on linkages to work only within the census areas and do not include places of employment outside the census areas.

The entries in this table show that the trip generation rate per household is higher for owners than renters with the exception of Kingston. The work trip linkage rate for owners varies from 1.04 in Kingston to 1.41 in Ottawa while the rate for renters varies from 0.90 in Thunder Bay to 1.12 in Sudbury and in Ottawa. The higher average rates in owner occupied

TABLE 31. Trip Generation Rates for Ontario Urban Areas Based on Tenancy

Census Area	Class	Linkages Produced	Number of Households	Trip Generation Rate
Guelph	Own	14,790	11,590	1.28
	Rent	6,720	6,625	1.01
Peterborough	Own	16,635	12,805	1.30
	Rent	5,610	6,005	0.93
Sarnia	Own	19,650	16,000	1.23
	Rent	6,675	6,635	1.01
Brantford	Own	19,755	16,550	1.19
	Rent	7,155	7,360	.97
Sault Ste. Marie	Own	20,235	14,935	1.35
	Rent	6,510	6,165	1.06
Kingston	Own	16,125	15,520	1.04
	Rent	13,020	12,400	1.05
Thunder Bay	Own	30,585	23,575	1.30
	Rent	7,770	8,615	0.90
Oshawa	Own	25,965	22,725	1.14
	Rent	10,320	11,255	0.91
Sudbury	Own	30,480	22,860	1.34
	Rent	18,570	16,535	1.12
Kitchener	Own	54,015	39,905	1.35
	Rent	30,105	26,645	1.11
Windsor	Own	61,335	52,240	1.17
	Rent	20,145	21,930	.92
London	Own	69,375	51,980	1.33
	Rent	37,545	35,155	1.07
St. Catharines	Own	77,805	63,895	1.22
	Rent	22,905	25,050	0.91
Hamilton	Own	114,510	93,140	1.23
	Rent	52,425	53,140	0.99
Ottawa	Own	120,780	85,595	1.41
	Rent	94,770	84,430	1.12

TABLE 32. Trip Generation Rates for Ontario Urban Areas Based on Period of Residence

Census Area	Class	Linkages Produced	Number of Households	Trip Generation Rate
Guelph	0-5 yrs.	11,115	9,540	1.19
	6+	10,410	8,665	1.20
Peterborough	0-5	9,765	8,650	1.13
	6+	12,180	10,120	1.20
Sarnia	0-5	12,180	10,720	1.14
	6+	14,160	11,935	1.19
Brantford	0-5	12,000	10,685	1.10
	6+	14,910	13,230	1.12
Sault Ste. Marie	0-5	11,430	9,520	1.20
	6+	15,300	11,570	1.32
Kingston	0-5	17,175	14,585	1.18
	6+	11,970	10,380	1.15
Thunder Bay	0-5	15,900	14,180	1.12
	6+	22,410	18,080	1.24
Oshawa	0-5	17,010	17,110	0.99
	6+	19,275	16,820	1.15
Sudbury	0-5	26,305	21,075	1.25
	6+	23,715	18,340	1.29
Kitchener	0-5	46,080	37,995	1.21
	6+	38,040	29,985	1.30
Windsor	0-5	36,060	32,940	1.09
	6+	45,505	41,145	1.11
London	0-5	57,780	47,795	1.21
	6+	49,155	39,360	1.25
St. Catharines	0-5	43,065	39,115	1.10
	6+	57,660	49,765	1.16
Hamilton	0-5	82,275	73,943	1.11
	6+	84,675	72,320	1.17
Ottawa	0-5	122,655	98,595	1.24
	6+	92,895	71,425	1.30

dwelling units are due to a variety of factors which might include larger household sizes with greater numbers in the labour force.

Table 32 summarizes the work trip generation rates for the fifteen urban areas based on period of residence categories. The generation rates for the longer period of residence are uniformly higher than the 0-5 year category except for Kingston. The work trip linkage rate varies from 1.11 in Windsor to 1.32 in Sault Ste. Marie for periods of residence 6 years and greater and from 0.99 in Oshawa to 1.25 in Sudbury for periods of residence of 5 years and less. The lower trip rates for the shorter periods of residency reflect no doubt the smaller household sizes of the recently established households.

Table 33 shows the mean trip lengths of households classified by tenancy type and of households classified by period of occupancy. This table shows that the mean trip length to work of renters is significantly different from that of owners with the biggest difference being in Ottawa where many work trips from owner occupied dwelling units are from residences outside of the greenbelt.

This table also shows that the average mean trip length to work from dwelling units that have been occupied for five years or less is in most cases longer than the average trip from households that have been occupied for six years or longer. This is reasonable since most of the new dwelling unit opportunities in urban areas are located on the periphery. The largest differences are in Windsor and Ottawa.

6.2 Travel Characteristics by Car Ownership Status

The most useful partitioned travel data available from the 1971 Census for transport planning purposes are the spatial interaction patterns

TABLE 33. Mean Trip Lengths for Socio-Economic Groups in Ontario Cities

	Mean Trip Length (km)			
	Owner	Renter	Occupancy < 5 Years	Occupancy 6+ Years
Guelph	3.6	3.8	3.9	3.5
Peterborough	3.3	3.0	3.3	3.2
Sarnia	6.4	5.7	6.4	6.0
Brantford	4.2	3.8	4.2	4.1
Sault Ste. Marie	4.6	4.1	4.5	4.5
Kingston	5.8	5.2	5.6	5.4
Thunder Bay	5.4	4.6	5.2	5.3
Oshawa	4.1	3.7	3.9	4.1
Sudbury	8.6	7.9	8.7	8.2
Kitchener	4.7	4.7	4.8	4.5
Windsor	7.7	6.7	7.9	7.1
London	6.6	5.9	6.6	6.1
St. Catharines	6.3	6.1	6.1	6.0
Hamilton	8.3	6.9	7.9	7.7
Ottawa	8.4	6.4	7.9	7.1

by car ownership status. The ability to estimate transit captivity and non-captivity from dwelling unit information is a very useful capability.

The regression analyses of work trip linkage formation rates reported in reference [1] demonstrated that the best prediction equations could be developed in terms of the dwelling unit composition of census tracts where the dwelling unit composition is expressed in terms of the number of single detached units and the number of attached dwelling units plus apartments. The use of the number of attached dwelling units as a third independent variable was found to yield statistically non-significant partial regression coefficients for many of the census areas. An initial analysis of the trip productions by the two car ownership groups using three independent variables demonstrated similar behaviour.

Tables 34 and 35 illustrate the prediction equations that have been developed to estimate the census tract amounts of captive and non-captive amounts of labour force, respectively, along with the statistical properties of these equations for the fifteen Ontario census areas. For the captive equations the partial regression coefficient of detached dwelling units varies from 0.156 in Kingston to 0.441 in Guelph and the t-statistics indicate that all of these coefficients are significant at the 1 percent level. The partial regression coefficient of attached dwelling units varies from 0.331 in Oshawa to 0.639 in Guelph. Pooled regression equations are also shown for each census area size class and for all census areas combined. However the F ratios are all significant indicating that the pooled equations cannot be used instead of the individual equations. Inspection of the two partial regression coefficients shows that labour force captivity generation from single attached dwelling units

TABLE 34. Census Zone Captivity Prediction Equations Using a Modified Dwelling Unit Composition

Census Area	α	b_1 (t_1)	b_2 (t_2)	R^2	F Ratio	DF
1. Guelph	-177	0.441 (5.38)	0.639 (5.72)	0.87		
2. Peterborough	13	0.299 (4.90)	0.421 (5.06)	0.81		
3. Sarnia	16	0.209 (11.10)	0.474 (13.14)	0.97		
4. Brantford	35	0.213 (5.03)	0.593 (9.41)	0.91		
5. Sault Ste. Marie	-104	0.287 (7.86)	0.705 (10.41)	0.95		
6. Kingston	-30	0.156 (5.32)	0.621 (14.76)	0.95		
GROUP 1-6	4	0.224 (12.62)	0.569 (21.06)	0.88		
CHOW TEST 1-6					3.121	15,72
7. Thunder Bay	10	0.273 (6.88)	0.555 (7.19)	0.94		
8. Oshawa	-133	0.340 (7.76)	0.331 (5.16)	0.87		
9. Sudbury	-40	0.275 (5.13)	0.562 (9.16)	0.86		
10. Kitchener	50	0.226 (2.80)	0.661 (7.62)	0.78		
GROUP 7-10	-94	0.300 (9.49)	0.605 (15.55)	0.85		
CHOW TEST 7-10					4.840	9,70
11. Windsor	-98	0.269 (6.51)	0.525 (9.95)	0.85		
12. London	17	0.255 (8.18)	0.521 (16.04)	0.94		

TABLE 34. (continued)

Census Area	α	b_1 (t_1)	b_2 (t_2)	R^2	F Ratio	DF
13. St. Catharines	9	0.178 (6.32)	0.537 (8.15)	0.93		
14. Hamilton	-38	0.265 (5.83)	0.536 (7.25)	0.71		
15. Ottawa	54	0.206 (5.40)	0.606 (18.91)	0.89		
GROUP 11-15	-13	0.219 (12.80)	0.589 (29.07)	0.86		
CHOW TEST 11-15					1.585	12,156
ALL CASES	-5	0.224 (19.82)	0.586 (42.44)	0.88		
CHOW TEST - ALL CASES					1.552	42,298

ZONE CAPTIVES = a + b_1 . Zone Single Detached Dwelling Units
+ b_2 . Zone Single Attached Dwelling Units and Apartments

and apartments is from two to three times greater than for single detached dwelling units.

The non-captivity equations presented in Table 35 show that the partial regression coefficient of detached dwelling units varies from 0.800 in Oshawa to 1.267 in Ottawa while the coefficient of attached dwelling units varies from 0.185 in Brantford to 0.542 in Sudbury. The attached dwelling unit partial regression coefficients are not significant for Guelph, Thunder Bay, Windsor and St. Catharines. The non-captive coefficients are two to three times larger for the detached dwelling units than the attached dwelling units. The table also illustrates that the pooled regression equations cannot be used in place of the equations for individual census areas.

The labour force participation rates vary between the Ontario census areas and a second set of regression equations have been developed for the case in which the labour force participation rates have been factored to a constant magnitude across all census areas. This operation allows the relative importance of car ownership status by dwelling unit type to be compared more directly across census areas. The prediction equations developed in this way for both captives and non-captives are summarized in Tables 36 and 37, respectively, along with their statistical properties. Pooled regression equations for all census areas are also shown in the two tables. The F ratio for the pooled captive equation is not significant indicating that the pooled equation may be used to represent the fifteen census areas. The F ratio for the non-captive pooled regression equation is just significant at the 1 percent level.

TABLE 35. Census Zone Non-Captivity Prediction Equations Using a Modified Dwelling Unit Composition

Census Area		b_1 (t_1)	b_2 (t_2)	R^2	F Ratio	DF
1. Guelph	143	0.838 (6.75)	0.319 (1.89)	0.83		
2. Peterborough	-12	0.989 (13.72)	0.262 (2.65)	0.94		
3. Sarnia	-79	1.073 (21.96)	0.328 (3.51)	0.98		
4. Brantford	18	0.953 (22.13)	0.185 (2.89)	0.98		
5. Sault Ste. Marie	86	1.023 (13.16)	0.353 (2.45)	0.94		
6. Kingston	32	1.185 (27.87)	0.247 (4.04)	0.98		
GROUP 1-6	-22	1.054 (37.72)	0.302 (7.09)	0.95		
CHOW TEST 1-6					3.364	15,72
7. Thunder Bay	-165	1.068 (13.89)	-0.008 (-0.05)	0.95		
8. Oshawa	146	0.800 (9.49)	0.415 (3.35)	0.88		
9. Sudbury	-94	1.076 (12.58)	0.542 (5.53)	0.91		
10. Kitchener	31	1.167 (10.42)	0.358 (2.97)	0.86		
GROUP 7-10	-69	0.982 (18.58)	0.558 (8.57)	0.87		
CHOW TEST 7-10					6.278	9,70
11. Windsor	67	0.938 (15.37)	0.129 (1.65)	0.90		
12. London	130	1.039 (12.69)	0.320 (3.74)	0.89		

TABLE 35. (continued)

Census Area	α	b_1	b_2	R^2	F Ratio	DF
13. St. Catharines	118	1.022 (17.87)	0.036 (0.27)	0.96		
14. Hamilton	-281	1.046 (19.19)	0.393 (4.43)	0.92		
15. Ottawa	-53	1.267 (16.06)	0.365 (5.52)	0.86		
GROUP 11-15	-7	1.017 (31.78)	0.384 (10.17)	0.87		
CHOW TEST 11-15					6.080	12,156
ALL CASES	-20	1.020 (49.64)	0.393 (15.64)	0.90		
CHOW TEST - ALL CASES					3.779	42,298

ZONE NON-CAPTIVES = $a + b_1$. Zone Single Detached Dwelling Units

$+ b_2$. Zone Single Attached Dwelling Units and Apartments

TABLE 36. Revised Captive Prediction Equations Using a Modified Dwelling Unit Composition

C(M)A	α	b_1 (t_1)	b_2 (t_2)	R^2
Guelph	-150	0.373 (5.38)	0.540 (5.72)	0.87
Peterborough	11	0.256 (4.90)	0.360 (5.06)	0.81
Sarnia	14	0.180 (11.10)	0.408 (13.14)	0.97
Brantford	31	0.190 (5.03)	0.527 (9.41)	0.91
Sault Ste. Marie	-82	0.227 (7.86)	0.556 (10.41)	0.95
Kingston	-26	0.134 (5.32)	0.532 (14.76)	0.95
Thunder Bay	8	0.230 (6.88)	0.467 (7.19)	0.94
Oshawa	-124	0.318 (7.76)	0.309 (5.16)	0.87
Sudbury	-32	0.216 (5.13)	0.443 (9.16)	0.86
Kitchener	40	0.183 (2.80)	0.534 (7.62)	0.78
Windsor	-89	0.244 (6.51)	0.477 (9.95)	0.85
London	14	0.208 (8.18)	0.424 (16.04)	0.94
St. Catharines	8	0.157 (6.32)	0.474 (8.15)	0.93
Hamilton	-33	0.232 (5.83)	0.469 (7.25)	0.71
Ottawa	42	0.163 (5.40)	0.478 (18.91)	0.89
ALL CASES	-5	0.200 (21.48)	0.472 (41.45)	0.88
CHOW TEST	F RATIO = 1.008		DF = 42,298	

Zone Captives/Aggregate Trip Production Rate = REVISED ZONE CAPTIVES

REVISED ZONE CAPTIVES = $a + b_1$. Zone Single Detached Dwelling Units

$+ b_2$. Zone Single Attached Dwelling Units and Apartments

TABLE 37. Revised Non-Captive Prediction Equations Using a Modified Dwelling Unit Composition

C(M)A	α	b_1 (t_1)	b_2 (t_2)	R^2
Guelph	121	0.709 (6.75)	0.270 (1.89)	0.83
Peterborough	-10	0.846 (13.72)	0.224 (2.65)	0.94
Sarnia	-68	0.923 (21.96)	0.282 (3.51)	0.98
Brantford	16	0.847 (22.13)	0.164 (2.89)	0.98
Sault Ste. Marie	68	0.807 (13.16)	0.278 (2.45)	0.94
Kingston	27	1.015 (27.87)	0.212 (4.04)	0.98
Thunder Bay	-138	0.899 (13.89)	-0.007 (-0.05)	0.95
Oshawa	137	0.749 (9.49)	0.388 (3.35)	0.88
Sudbury	-74	0.848 (12.58)	0.427 (5.53)	0.91
Kitchener	25	0.943 (10.42)	0.289 (2.97)	0.86
Windsor	61	0.852 (15.37)	0.117 (1.65)	0.90
London	106	0.847 (12.69)	0.260 (3.74)	0.89
St. Catharines	104	0.902 (17.87)	0.032 (0.27)	0.96
Hamilton	-246	0.917 (19.19)	0.344 (4.43)	0.92
Ottawa	-41	0.999 (16.06)	0.288 (5.52)	0.86
ALL CASES	-22	0.882 (57.08)	0.301 (15.94)	0.92
CHOW TEST	F RATIO = 1.685		DF = 42,298	

Zone Non-Captives/Aggregate Trip Production Rate = REVISED ZONE NON-CAPTIVES

REVISED ZONE NON-CAPTIVES = a + b_1 . Zone Single Detached Dwelling Units

+ b_2 . Zone Single Attached Dwelling Units and Apartments

Inspection of the captive partial regression coefficients of detached dwelling units shows that they vary from 0.134 in Kingston to 0.373 in Guelph and those of attached dwelling units from 0.309 in Oshawa to 0.556 in Sault Ste. Marie. The partial regression coefficients for the pooled equation indicate that for average labour force participation rates the captive generation rate is 2.3 larger for attached dwelling units than detached dwelling units.

Inspection of the non-captive partial regression coefficients of detached dwelling units shows that they vary from 0.709 in Guelph to 1.015 in Kingston and for attached dwelling units from 0.104 in Brantford to 0.427 in Sudbury with the coefficients for Guelph, Thunder Bay, Windsor and St. Catharines being non-significant. The pooled equation suggests that the non-captive coefficient of detached dwelling units is about 3 times larger than for the attached dwelling units.

Table 38 summarizes the mean home to work transport network distances for non-car owners and car owners for the fifteen Ontario census areas. This table shows that there are some large differences between the mean trip lengths for the two groups in some areas with the largest differences being in Hamilton, Sudbury, Ottawa, Sarnia and Windsor. The mean values presented in this table will differ from those presented earlier since all trips are included in the averages calculated in Table 38 and not just the linkages between official census tracts. Some of the spatial differences in behaviour of the two groups have been presented in Chapter 5.

TABLE 38. Mean Trip Lengths for Non-Car Owners and Car Owners in Ontario Census Areas

Census Area	Mean Trip Length (km)	
	Non-Car Owners	Car Owners
Guelph	3.4	3.9
Peterborough	3.0	3.4
Sarnia	4.5	6.7
Brantford	3.5	4.4
Sault Ste. Marie	3.5	4.9
Kingston	4.3	6.0
Thunder Bay	4.6	5.6
Oshawa	3.9	4.1
Sudbury	6.4	9.2
Kitchener	4.1	5.1
Windsor	6.0	8.0
London	5.1	6.9
St. Catharines	4.7	6.8
Hamilton	5.8	8.8
Ottawa	6.1	8.3

CHAPTER 7

SOME GENERALIZATIONS ABOUT WORK TRIP MODELLING

The overall objective of the studies of the 1971 census journey to work data described in reference [1] and in this report was to understand better the home to work spatial linkages patterns in the Ontario census areas. The basic motivation was the idea of developing generalized relationships across all urban areas that could be used in transport planning studies. A traditional route has been followed in these studies consisting of the three components of (i) trip production analysis, (ii) trip attraction analysis, and (iii) trip distribution analysis. The hypothesis implicit in this approach was that the parameters produced by these three sub-analyses for each census area could be related to some broad properties of the census areas such as population size and employment sector composition. This chapter examines the extent to which these overall objectives have been met and outlines some of the implications for urban transport modelling.

7.1 Trip Generation Analyses

The trip generation analyses described in this report and the earlier study have demonstrated very clearly that high quality prediction equations for estimating the labour force composition of census tracts may be developed in terms of the dwelling unit composition of census tracts. The important feature of these prediction equations is that the partial regression coefficients are consistent across all of the

Ontario census areas with any real differences being due to variations in the labour force participation rates across census areas. These regression analyses have also shown that high quality prediction equations may also be developed in terms of census tract dwelling unit composition for estimating labour force transit captivity and non - captivity. Once again when differences in the labour force participation rates between census areas are removed from the data the partial regression coefficients are very consistent across all census areas.

These findings would suggest that work trip generation studies be standardized in terms of dwelling unit composition for all urban transport studies subsidized by the Ontario Ministry of Transportation and Communications. The dwelling unit composition and distribution in urban areas are perhaps the highest quality land use data available. More importantly it is a variable which is influenced strongly by municipal planning policies and may be reliably estimated for future time horizons. Dwelling units are also the central item in municipal assessment records and the opportunities for developing more sophisticated dwelling unit classification schemes and relating travel behaviour to dwelling unit classes are enormous. Assessment records also contain information on household structure which would allow reliable correlations to be established between dwelling unit type and household structure variables such as labour force composition.

Trip attraction equations cannot be established from the census data since information on the characteristics of places of

employment is not collected. However, if a common system of employment and land use classification could be established across all Ontario municipalities then this trip attraction information could be exploited fully. An unsuccessful attempt was made to establish relationships between captive and non-captive trip attractions and employment sector type. Significant transport planning benefits would be derived from the adoption of a Province-wide municipal land use classification scheme and this may gradually emerge from the current studies of assessment practices.

7.2 Trip Distribution Modelling

The studies of trip distribution across the Ontario census areas have demonstrated that cross-sectional type gravity models are very poor in explaining the home to work patterns observed in the 1971 census. A large number of modifications to the basic gravity model have been explored in this project and these have produced only marginal improvements to the explanatory power of the gravity model. However these analyses of the gravity model have provided important insights into its behaviour.

The analyses which have been conducted have demonstrated very clearly that the best form of the gravity model to use is the doubly constrained version. However, the trip tables calculated by the gravity model when compared with the observed trip table exhibited goodness of fit levels equivalent to about a 75 percent randomly introduced error range. A detailed analysis of the calibrated doubly constrained models showed that the model properties are dominated almost completely by the

spatial distributions of trip ends and that the so-called behavioural parameter of the deterrence function has only marginal effects.

While the doubly constrained versions of the gravity model possess the best goodness of fit characteristics it is difficult to develop any rational interpretations of the production end and attraction end balancing factors or the deterrence function behavioural parameters either within or between census areas. Balancing factor magnitudes do not vary in any systematic way within urban areas and the deterrence parameter magnitudes have only a very weak association with mean trip length between urban areas. If the calibration parameters were rational then they should be interpretable in some way both within and between urban areas.

Comparisons of the qualities of models calibrated in terms of network distances and network travel times showed that distance is the best measure to use. In very large urban areas the use of network times improved the goodness of fit of some of the outlying zones but the distance-based models were superior. Once again this observation reinforces the previous conclusion that doubly constrained model behaviour is dominated by the structural parameters and not the behavioural parameter.

The strategy adopted in this research project for improving the goodness of fit of gravity models was to stratify trip ends and to distribute trip interchanges within each of these strata and then to sum the trip tables across strata. These attempts resulted only in marginal improvements to the doubly constrained models but there were

some significant improvements to the production constrained models.

Perhaps the greatest opportunity for improving the quality of gravity models is to stratify by period of development and to develop effectively a dynamic trip distribution model. The cluster analyses showed very clearly the importance of the staging of development on the evolution of commuting sub-regions. This strategy for improvement could not be implemented since information on the period of development of employment opportunities is not available from the census information.

These studies of trip distribution across the fifteen Ontario census areas have demonstrated quite clearly that there is little point in calibrating and using the traditional cross-sectional type gravity models. The model-estimated base year trip tables are so different from the observed trip tables that any trip tables estimated for the horizon year are of little use. Time-staged or dynamic forms of the gravity model must be developed if estimates of future travel are to have any credibility. All data collection exercises should record the period of location at both the home and work ends in order that trip tables partitioned by period of development may be developed. It is also essential that every effort be made to collect journey to work data in the 1981 census in order to develop a consistent set of time series data on home to work linkages.

CONCLUSIONS AND RECOMMENDATIONS

1. Consistent and high quality prediction equations for estimating the labour force composition of census tracts have been developed for all Ontario census areas in terms of the dwelling unit composition of census tracts.
2. Similar prediction equations have been developed for estimating the transit captivity and non-captivity of members of the residential labour force.
3. Equivalent work trip attraction equations cannot be established since independent measures of land use at the employment-end are not available.
4. It is recommended that work trip generation studies be standardized in terms of dwelling unit composition. Centralized assessment records should be used to establish Ontario-wide dwelling unit classifications and consistent employment-end land use classifications.
5. Bi-proportional matrix balancing techniques may be used to establish the interaction structure with urban areas that is independent of zone size effects. Cluster analysis may be used very effectively to establish the commuting structure of urban areas and 1971 commuter-sheds have been established on a consistent basis for the Ontario census areas.
6. Additional empirical analyses of other census data are required to establish more clearly the determinant of these commuter-sheds in the Ontario census areas.

7. The goodness of fit statistics such as the coefficient of determination and chi-squared are inadequate for assessing the qualities of calibrated gravity models. The best statistic to use for comparing observed and estimated trip matrices is the phi-statistic of information theory.
8. Cross-sectional gravity models are very poor in explaining the observed journey to work patterns in Ontario census areas having goodness of fit characteristics similar to a trip table created by introducing random errors of 75 to 100 percent in the observed trip interchange magnitudes.
9. Doubly constrained gravity models have superior goodness of fit characteristics to the singly constrained models. With the doubly constrained models compliance with the trip end constraints is the major determinant of the estimated trip interchange magnitudes with the deterrence function having the secondary effect. The contribution of the deterrence function increases with urban area population.
10. Gravity models stratified by geographic sub-region or by socio-economic group result only in marginal increases in the goodness of fit. The largest increase in goodness of fit was achieved by stratifying trip makers into car owners and non-car owners.
11. The most important determinant of commuting that cannot be captured through trip end stratification and constraint is the timing of development. Longitudinal models would be required to reflect timing of development effects in which trip interchange elements are partitioned into different sets to reflect different periods

of development.

12. Network distances are adequate for use as a surrogate of generalized travel costs in smaller and medium sized urban areas. Network travel times result in some improvements in the goodness of fit of gravity models in the largest cities.

13. Because of the poor goodness of fit characteristics of the traditionally used gravity model it is recommended that their use as an estimator of future trip distribution patterns be discontinued in urban transport planning studies. It is recommended that the traditional cross-sectional model be replaced by a longitudinal form of the gravity model. Some development work is required in order to implement this recommendation.

REFERENCES

[1] Hutchinson, B. G., The Journey to Work in Urban Ontario, Volume 1: Data Analyses and Interpretation, Final Report OJT and CRP Project No. 8507, Ministry of Transportation and Communication, Ontario, 1978.

[2] Bacharach, M., Biproportional Matrices and Input-Output Change, Cambridge University Press, London, 1970.

[3] Furness, K. P., Time Function Iteration, Traffic Engineering and Control, November, 1965, pp. 458-460.

[4] Williams, W. T., and G. N. Lance, 'Hierarchical Classificatory Methods', in Enslein K., A. Ralson, and H. S. Wilf (Editors), Statistical Methods for Digital Computers, Wiley Interscience, New York, 1977.

[5] Baxter, R. and I. Williams, An Automatically Calibrated Urban Model, Environment and Planning A, Vol. 7, No. 1, pp. 3-20, 1978.